

Dangers of disappointment at Rio

Global environmental problems have dictated the need that next week's conference at Rio should be as much concerned with development as with global warming. Both deserve more than a couple of weeks.

THE great UN conference on Environment and Development next week at Rio de Janeiro is bound to be a disappointment. Too much is too widely expected of it. Those who have christened the event the "Earth Summit" look to Rio for a once-and-for-all resolution of global problems; already they are complaining that the draft treaty on global warming has been unacceptably "watered down" (see *Nature* 357, 97; 14 May 1992) and, more to the point, that neither the declaration the conference will adopt nor the discussion that precedes it is likely to deal with the rapid population growth that makes a nonsense of the aspirations of the poorest countries and of their well-wishers (see *Nature* 357 177; 21 May 1992). It is in everybody's interest that those banging the drum for Rio should discount their inevitable disappointment in advance. Otherwise, disappointment will turn to anger and rational discussion of important problems will be corrupted by consequent incoherence. That is why the origins of the conference should not be forgotten. Rio is happening because of the clamour, in the mid-1980s, that development projects in poor countries should qualify for multilateral technical and financial assistance only if they were environmentally sound. The argument had a certain cogency. Why should rich countries finance the building of paper-pulp plants on the banks of virgin lakes in developing countries while energetically restricting the operations of their domestic plants? More generally, can it make sense to export sources of pollution from countries rich enough to purchase freedom from pollution to those so poor that they have no choice but to put up with it? Especially when some forms of pollution are not confined by the frontiers of the recipient states?

Connections

The snag is that the sensible course of action may slow the pace of economic development in poor countries, which may then reasonably ask to be compensated for the restraint forced on them by others. That there is thus a connection between environment and development is self-evident. It made sense that there should be an international conference to hammer out an understanding between rich and poor countries on the terms on which the latter should be compensated for adopting the prescriptions of the former. No such understanding would have been possible without agreement on mutually acceptable environmental goals.

Global warming

That agenda has been overtaken by events, most obviously

by the recognition that global warming consequent on the greenhouse effect may become, in some decades, a serious threat to people's comfort and prosperity, even survival. The Intergovernmental Panel on Climate Change (IPCC), which in 1990 dramatized that prospect, has ensured for global warming a place on the agenda at Rio. That is as it should be. Moreover, the Vienna Treaty on the protection of the ozone layer, and the Montreal Protocol thereto that will in due course restrict and eliminate chlorofluorohydrocarbons from the atmosphere, show that an agreement to head off global warming is within the bounds of what is diplomatically possible. Rio is therefore a splendid opportunity to put flesh on the bones of that hope.

But too much flesh at this stage could be a hindrance, not a help. This journal has consistently argued, perhaps *ad nauseam*, that Rio should ideally produce the framework of an international agreement, a binding recognition in the form of a treaty whose signatories acknowledge that global warming is a serious common threat and that they will take steps to fend it off when it is plain what action will be best. The case for reaching the framework of an understanding now is what it has been for the past five years — that the timescale of climate change is comparable with the time that will be needed to negotiate a more specific agreement, complete with numbers specifying what gases may be emitted by whom, in what circumstances and when.

None of that implies that global warming is not already a serious matter. On the contrary, there is every reason why increased surface temperatures will follow increased greenhouse gas concentrations as surely as effects usually follow causes. The important article on page 293, an analysis of the physical consequences of the improved estimates of emission patterns just published by IPCC, is a striking illustration of how predictions of global warming hang more coherently together as further refinements of atmospheric mechanisms are taken into account. (Elsewhere in this issue, on pages 315, 318, 320 and 322, there are other articles showing how rich the problem of global warming remains.) The same article shows that while the certainty of global warming can only be avoided with the greatest difficulty, its magnitude (and thus timing) remains hugely uncertain. Crudely, plausible estimates of the increase in surface temperature between now and some point in the future lie within a range whose upper limit is three times greater than the lower, but, with likely greenhouse gas emissions, both the lower and the upper limits are positive.



Framework

So would not a Rio treaty with numbers restricting greenhouse gas emissions be preferable to a mere framework? That is what the enthusiasts have been asking for. They are wrong. For one thing, there are many countries that would not sign (the United States, for example) while there are others that would, but whose signatures could not carry much more weight than the ink with which they are written. But even when a large group of countries, say those of western Europe, are willing collectively to restore greenhouse emissions at the end of the decade to the levels of 1990, it is not self-evident that such a decision can be the wisest. At worst, it may relieve others of what should be a mutual responsibility. But self-denial, probably deflationary in its economic effects, may also get rid of wealth that, for the greater good, should better be spent on causes other than global warming.

One of the serious flaws of Rio is that it presents important and exceedingly complicated questions in absolute terms. But is not global warming a threat to the continued habitation of the surface of the Earth by people? Of course. But so (still) are the proliferation of nuclear weapons, the risks of unknown pestilences, the gaps between rich and poor within nations as well as internationally and the pace of the world's population growth (which can only make those gaps more dangerous and offensive). It is one thing to accept that global warming is a threat to be reckoned with, and another to pretend (as some do) that every action, however symbolic, that may partially contribute to its abatement must take precedence over every other wise course. What if, for example, money spent on improved child-health in poor countries, by blunting the advantages of large families, did more to reduce the potential demand for fossil fuels than the same sum spent on reducing rich countries' use of oil?

Mechanism

What the cause of global warming needs is a durable mechanism for redefining the goals for abating global warming as the decades ahead tick by, thus specifying protocols that may be added to the kind of framework treaty that Rio will produce. With the best will in the world, it will not suffice that a bunch of climatologists such as IPCC should be constituted as a committee to recommend what action should be taken. The practical problems of abating global warming are too subtle. First, there are problems of equity as between rich and poor (and fuel-rich and fuel-poor) countries; on what basis do the rich compensate the poor for their restraint? Then there are problems of compliance; how is cheating detected, what penalties does it entail and how are they administered? Questions such as national maladministration, or the pace of population growth, should strictly be parts of the same web of imponderables; how will they be brought in against the protests of those who insist that national sovereignty is paramount?

The conference at Rio should give more than a little thought to the progressive disentangling of these questions over the next several decades. It is not the stuff with which *ad hoc* committees, however expert, can cope. The need is for some permanent institution of thoughtful and probably scholarly people who can plan and execute the research

leading to the understanding of the whole world that the problem of global warming has occasioned. (The United Nations could do worse than subvert the International Institute of Applied Systems Analysis in Vienna, now heading for the doldrums, for such a purpose.) That would not obviate the need for regular negotiating sessions to agree further protocols, but it could ensure that one agenda began where the previous one left off, and that the agenda items themselves were sensible.

Priorities

Not the least of the tasks for the post-Rio period is that of putting global environmental goods in some order of priority. The abatement of global warming would come near the top of any simple list, but a simple list cannot be meaningful. As many are fond of saying, everything hangs together. The issue of biodiversity, likely to be given a good hearing at Rio, also needs to be put carefully in its place. The argument, which owes as much to the doctrine of sustainable development as to a present emergency, is that species extinction impoverishes the surface of the Earth, perhaps (in the long run) dangerously so. At the extreme, the cause is held to justify the conservation of all habitats, natural and otherwise, in their present state, if necessary by the transfer of funds to poor countries. But the basis of understanding necessary to strike a fair price is mostly, at present, inadequate. And who, now, can be sure that the funds transferred will not be needed for the abatement of global warming?

Another flaw in the constitution of the Rio conference is that its implicit (but not hidden) agenda is the solution of all the world's problems, from those of poverty and its erosion of human dignity to the emission of greenhouse gases (often in the interests of abating poverty) and the risk that they will cause a catastrophic slumping of the West Antarctic ice-shelf. People like Mr Maurice Strong, the Canadian chairman of the conference, are asking for what seem like snap solutions of these great issues. They are right to imply that global warming, eminently an international problem, demands global solutions. So, for that matter, does the plight of poor countries (which is why institutions such as the World Bank exist). Most aid agencies and some governments have been saying as much for years. If Rio can wring from governments in general a commitment to those principles, and a resolve deliberately to act upon them, that will be an invaluable. If the issue of population growth is ducked — there will have to be another meeting. Of its nature, the meeting in Rio is not one at which permanently valid bargains can be struck to compensate poor countries for promised restraint. Talk, over the past few months, of a compensation fund of more than \$100 million a year has inevitably raised expectations among potential recipients, just as the expectations of environmentalists converging on Rio have been raised by talk of action this day (or at the latest, tomorrow) on greenhouse emissions and even the conservation of still uncatalogued species. Those expectations are fated to be disappointed at Rio. The hope must be that the disappointed will make the long-term commitment to the good causes advertised at Rio. □

Congress compromises on rules governing NIH policies

Washington. Under the cloud of a threatened presidential veto, Congress last week agreed on a compromise package of legislation affecting the National Institutes of Health (NIH) that moderate many of its most controversial incursions into biomedical research politics. Although public attention on the bill, known as the "NIH Revitalization Amendments of 1992", has been focused on its attempt to overturn the current federal ban on fetal tissue research (see sidebar), scientific groups have been equally concerned about its language on a variety of subjects from women's health research to conflict of interest. Until last week, the House of Representatives draft defined scientific misconduct and conflict of interest in research and prescribed detailed procedures to prevent it.

In a conference last week, the House and Senate agreed on a compromise bill on which they are expected to vote this week. Administration officials say that President George Bush intends to veto the bill, both for its fetal tissue provisions and for a host of other clauses that he considers a pre-emption of the administration's own policy-making process. But Congressional supporters of the bill are optimistic that they will be able to overturn a veto. Among its provisions, the joint bill would:

- Forbid the Secretary of the Department of Health and Human Services (HHS) from arbitrarily halting research on ethical grounds, as long as it has been peer-reviewed and approved by an Institutional Review Board. To withhold funding for such a project, the HHS secretary would have to get the approval of an independent ethics advisory board. On the other hand, the secretary would also have to demonstrate that any surveys of sexual behaviour (such as the three the administration has cancelled under political pressure in recent years) will help reduce sexually transmitted disease or otherwise improve health. To placate conservative Senator Jesse Helms (Republican, North Carolina), the bill bans two sex surveys that have been political footballs for years.

- Establish the Office of Scientific Integrity as an independent entity within HHS. This would remove it from NIH, which NIH director Bernadine Healy has also recommended.

- Require HHS to develop a new definition of scientific misconduct. In its own bill, the House had actually supplied a definition, but researchers attacked it as impossibly broad. The conference report instead requires HHS to set up an independent "Commission on Scientific Integrity" to advise

HHS on the definition and other aspects of policing scientific misconduct. A section that would have forced journals to retract research that had been shown in error (see *Nature* 357, 2 & 7; 7 May 1992) was deleted in the conference agreement.

- Require regulations defining and protecting against financial conflict of interest in NIH-funded research. NIH has been working on such regulations for several years; Healy says the draft regulations now being prepared for release are essentially in line with the congressional requirement. Contrary to earlier versions of the bill, the conference agreement would allow institutions to develop their own procedures to comply with the NIH regulations.

- Ensure that NIH has regulations protecting whistleblowers from retaliation. In practice, this would mean that institutions receiving NIH funding would have to put in place guidelines to protect those who had made scientific misconduct allegations "in good faith" and cooperated with investigators.

- Establish criminal penalties for activists who attack animal facilities. This would include anyone who took or released animals, stole documents or damaged the facilities. Punishment would include fines, imprisonment and restitution. However, the bill would also require NIH to develop a

plan to minimize its use of animals and to conduct research on animal alternatives.

Research groups last week seemed generally relieved at the final language. "If it had been up to us, it would have been more like 15 lines than 30 pages", says David Moore, a congressional lobbyist for the American Association of Medical Colleges. "But overall, it's not a bad compromise." Healy, however, attacked the bill as both unnecessary and intrusive. "We don't think we need to be reauthorized", she says. The administration "has a number of concerns with the bill. Whether it was the fetal tissue part that pushed them over the top [to a veto], I don't know." Nevertheless, Healy says that NIH can live with the consequences should Congress collect enough votes to overturn the expected veto and turn the legislation into law.

"On issues like women's health and minority health, we've already moved ahead of the bill", she says. Since the legislation has been in the making for more than three years, NIH has already had ample opportunity to respond to many of the congressional concerns that are behind the language. Many of the other provisions in the compromise language, Healy says, either reflect moves already afoot at NIH or extend current programmes.

Christopher Anderson

Researchers reject fetal tissue banks

Washington. In an eleventh-hour bid to derail congressional legislation overturning his ban on research involving the transplantation of fetal tissue, President George Bush promised last week to set up a "fetal tissue bank" to collect material from miscarriages and ectopic pregnancies — the only sources of fetal tissue that avoids the political minefield of the abortion issue.

But research and health groups denounced the move as a "cruel hoax". They said that little tissue from such fetuses would be usable for medical purposes or research and that viable tissue would be nearly impossible to obtain. "It simply would never work", Eugene Redmond, director of Yale University's fetal tissue transplant programme, wrote in a letter to Congress last week.

In explaining the tissue bank proposal, James Mason, assistant secretary of the Department of Health and Human Services, estimated that of the 200,000 or so spontaneous abortions that occur in US hospitals each year, about 2,000 could be used in a tissue bank for research. But Redmond cited a study of fetal tissue from spontaneous abortions in three New York hospitals to argue that such usable abortions occur, on the average, only 1.4 times a year in each hospital. A team of three or four specialists need to be on hand to dissect the brain region in the fetus. "We would have to station such a team on call at each hospital at any time a pregnant woman came in with a threatened miscarriage," he wrote. Spontaneous abortions usually occur for a reason that renders the tissue unusable for research or transplant.

Redmond also rejected the White House suggestion that, once a bank is established, the fetal tissue could be perpetuated and grown in cell lines. Noting that such a process would require new fundamental knowledge of molecular and genetic mechanisms, he emphasized that "these techniques do not exist and cannot be legislated into existence".

C.A.

Max Planck supports research in east as cost of reunification climbs

Halle. It is now clear that scientists in western Germany will have to pay part of the cost of reintegrating their scientific colleagues from the east. But recent decisions by the Max Planck Society to fund more than 30 research groups in the former East Germany, including four new research institutes, has begun to restore confidence that the research enterprise in the new Germany will be strengthened by reunification.

Despite earlier predictions that research in East Germany might be too low in quality to be worth saving, detailed evaluations of eastern research by the society and by the German research advisory council (the Wissenschaftsrat) have highlighted pockets of research excellence in the east. Reversing its previous policy, the society has decided to fund two new physics institutes and two for biological research when money is available. Many smaller projects in universities are already being funded.

The Halle Institute for Solid State Physics and Electron microscopy has become the first Max Planck Institute in the former DDR. It survived a restructuring. The Max Planck Society, which funds 56 large basic science research institutes in western Germany, was originally dubious of bailing out eastern research. It was concerned that lingering attitudes of the communist regime might somehow 'infect' its own research philosophy. Pessimism had been fuelled by attempts to evaluate East German science before the Berlin Wall came down in 1989. Obscured by Iron-Curtain secrecy, these had indicated that its level might be no higher than Albania's.

But prejudices began to fall away when

early results of the detailed evaluations commissioned by the research advisory council in summer 1990 highlighted pockets of research excellence in the east. This coincided with the appointment in June 1990 as president of the society of Hans Zacher. Zacher is more sympathetic than his predecessor to the plight of the east.

With a more optimistic outlook, the society set up its own evaluation committees and, in cooperation with the research advisory council, identified the areas of research it wished to support. By the end of last year, these included 28 individual groups in universities whose work will be supported financially for five years (see sidebar) and two institutes — the Halle institute reconstituted in January and the Institute for Colloid and Interface Research, to be established within the next few years at a location yet to be decided.

The former Institute of Solid State Physics and Electron microscopy in Halle pioneered the electron microscopic decoration technique which allows the properties of crystal structures to be correlated with their defects. And with its light still well hidden under a bushel, the institute was also for many years ahead of the world in the development of electron microscopes. According to Johannes Heydenreich, vice-director of the institute since 1962 and now one of three in-house directors of the newly created institute, success was due to the working environment created by its first director, Heinz Bethge, who retired in 1985. "He created a unique atmosphere of support and cooperation", says Heydenreich. "He had a classically open-minded approach to research.

And, unusually for the DDR, if a senior staff member was invited to a conference in the West, an infrequent enough event in itself, he insisted that one of the younger scientists go along."

But a certain price was exacted for this success. Like all research institutes in East Germany, the Halle institute was forced to spend at least half its time solving industrial problems. And each institute had to generate a certain proportion of its funding from industrial sources. In this way, the state ensured a firm and very direct relationship between fundamental and applied research.

Pressure to raise industrial income intensified as the economy of the DDR began to collapse in the 1980s, so that, by 1989, eastern institutes were raising half their running costs this way. The successful Halle institute was even returning a surplus to the state.

But such direct support for industry had no place in western research plans, so that evaluation committees set about identifying the redundancies eastern scientists had been fearing. In the event, most institutes have shed be-



**Research Minister
Heinz Riesenhuber**

tween a third and one half of their staffs, but Halle is again exceptional.

Max Planck took on 100 staff (including 45 scientists), although one-quarter were hired on three-year contracts. A group of 15 are now sponsored by Freiburg's Fraunhofer Institute, which specializes in applied research in microstructure technology. Five scientists have set up an independent micro diagnostic business with temporary assistance from the German Research Council (DFG). Nine are working on short-term grants from the German research councils and the rest left voluntarily for jobs in the west.

Heydenreich, whose lack of political influence with the old regime had caused him to be passed as Bethge's successor, was made the first of three directors of the new institute. He will be joined by two West Germans: Jürgen Kirschner from the Free University in Berlin and Ulrich Gösele from Duke University in North Carolina, whose interest in conducting research at Halle has

Social sciences found worth funding

Munich. Surprisingly strong research in the social sciences by scientists in eastern Germany has forced the government to examine how the humanities should be funded in a unified Germany.

The finding that the former DDR was doing good work in subjects ranging from linguistics to contemporary history and history of science came as West Germany was debating how to strengthen its own, long-neglected humanities programmes. To solve both problems, the German research advisory council (the Wissenschaftsrat) appointed the Max Planck Society as trustee to seven identified groups in the East whose research was recognized as being of high quality. The committee appointed to oversee the centres will make recommendations on how the humanities might best be fostered in the future in a unified Germany.

Hans Zacher, president of the society, does not expect the committee to support the research advisory council's original suggestion to set up a society for social sciences analogous to the Max Planck, which directs centres for science research and receives half its funding from the federal government and half from regional governments. "It's more likely that each group will be affiliated to a university but receive special funding from an organization like the German research council (DFG)", he says.

A.A.

outweighed the disadvantages of a lower standard of living.

Setting up the Halle institute has so far cost the Max Planck Society DM9.8 million (US\$6 million). Half of this money is provided by the federal government and the other half by the eastern states. The funding is independent of Max Planck's 1992 DM1,400 million budget, split equally between regional governments in the west and the federal government. In some respects the Halle institute was added to the fold on the cheap. Salaries are only 60 per cent of those in the west, and it will be a few years before they will be equalized. Although pay is still better than under the DDR, the cost of living is rising rapidly in the east and the staff are keen for wage parity. Last week, Bonn announced that rents (which are state controlled in Germany) in eastern states are to be doubled by the end of the year. The second Max Planck institute planned in eastern Germany will be of a similar size to Halle. Three research groups, from Adlershof and Teltow near Berlin and from Freiberg in Saxony, will be brought together to work on interdisciplinary projects investigating the structure, dynamics and physical and chemical properties of colloidal systems. Its location will be decided this summer.

The society is also keen to safeguard two further areas of research excellence in the former DDR. In March it announced its intention to support two further institutes — one in infection biology and the other in plant molecular physiology — as soon as money becomes available.

But when will that be? And will the west accept the increasingly high price of supporting the political ideals of those who wish to raise east German standards to their own? In the sudden political and economic uncertainty of Germany, these questions are difficult to answer.

Despite the special funds that have been set up to support research in the former DDR, it is clear that the government's persistent denial that the west must make sacrifices has lost credence. In fact, last week Research Minister Heinz Riesenhuber announced that 100 jobs would be lost in large research centres in the west so that money can be redirected "to bring the level of research in East Germany to levels in West Germany as quickly as possible." It was the government's first admission that there will be scientific losers as well as winners in its effort to reunite the formerly divided country. And like the DFG, the 5 per cent annual budget increase to which the society is committed until 1995 will be swallowed up by staff costs following the inflationary public sector wage settlement earlier this month (see *Nature* 357, 182; 1992).

The spirit of Germany may be willing, but it is becoming increasingly difficult to maintain the extraordinary speed and efficiency that has marked the first few years of reintegration of East German science.

Alison Abbott

Max Planck funds small projects

Munich. In the former DDR, research was mainly conducted in the 60 research units run by the National Academy of Science and concentrated around Berlin. Relatively little activity took place at universities, the traditional home for innovative research in West Germany.

But this situation is changing. The government's decision in June 1990 to harmonize research policy prompted the Max Planck Society, which normally sponsors only large-scale research that cannot be conducted within universities, to offer temporary support to deserving groups.

To protect research identified as very strong by evaluation committees of the society and the German Research Council (DFG), and to strengthen future university research, Max Planck is funding and administering 28 research groups in areas ranging from physics, to biochemistry to ecology. Funding will be restricted to five years. If the research develops well during this time, the universities will be expected to take over the financing directly.

The jewel in the eastern crown is Gunther Fischer's group from Halle. In the late 1980s, Fischer was the first to identify and purify *cis-trans* peptidylprolyl isomerase, an enzyme which catalyses a rate-limiting step in protein folding. This enzyme was subsequently shown to be identical to cyclophilin, which binds the immunosuppressant drug cyclosporin, and it has helped to open up an important new field of biochemical immunology.

A special fund for these projects has been provided by the federal and regional governments, who normally split the costs of running the Max Planck Society. But the 1992 allocation of only DM130,000 (US\$80,000) per researcher, amounting to DM98,500 million in total, has been criticized as inadequate by Hans Zacher, president of Max Planck. Zacher would like to raise the average grant to DM180,000 in 1993.A.A.

Researchers look beyond UNCED

London. As diplomats, politicians and representatives from non-governmental organizations gather next week in Rio de Janeiro for the United Nations Conference on Environment and Development (UNCED), the environmental researchers whose work set the agenda for what is popularly known as the Earth Summit are concerned about what will happen afterwards.

"The critical period is what follows UNCED, and scientists are concerned about maintaining whatever momentum the conference has generated," said Jim Dooze, president-elect of the International Council of Scientific Unions (ICSU). "The key thing, post-UNCED, is science forging new partnerships with business, between the natural sciences and social sciences, and with policy makers."

Although some researchers bemoan the lack of scientific input into the conference itself, most recognize that Rio will be crowded enough over the next three weeks without them. The discussions that will take place will be largely political — UNCED is not now a scientific conference, if it ever was — and more than 100 heads of state are expected to address the delegates.

Although science may have been absent from the proceedings, environmental researchers have had an impact on the substantial issues being addressed. Bodies such as the Intergovernmental Panel on Climate Change and ICSU's ASCEND-21 meeting

(An Agenda of Science for Environment and Development into the 21st Century) have submitted weighty documents to the conference, as have individual scientific groups. Last week, for instance, the University of Oxford's Environmental Change Unit published a report entitled "Climate Change and World Food Supply". In addition, the efforts of researchers are represented in the various texts to be considered by UNCED delegates, and even if they are not adopted, it is hoped that their contribution will be recognized.

"Science has had a considerable input", says Tim O'Riordan, of the Centre for Social and Economic Research in the Global Environment, University of East Anglia, "but it is a matter of whether UNCED understands that, and if the documents produced from the conference reflect that. If that is not the case, then science will be damaged."

One way to bring that about is for Maurice Strong, secretary general of UNCED, to set up an office at the UN to set the agenda for future scientific work, including a roadmap that points out the gaps in existing knowledge and to areas that are most important. Although scientific programmes will undoubtedly continue in the wake of UNCED, researchers believe that the conference itself will heighten awareness of the issues and put pressure on governments to increase their monitoring of environmental problems.

Ian Mundell

NIH defends gene patents as filing deadline approaches

Washington. With an eye on a 20 June deadline to extend its controversial cDNA patents to the rest of the world, the US National Institutes of Health (NIH) last week vigorously defended its approach in spite of the concerns of its government partner in the US genome project and criticism from other countries.

Speaking at a public meeting of an interagency committee that is formulating a US policy on gene patents, Bernadine Healy, the NIH director, repeated that NIH was following both legal precedent and a legislative mandate to commercialize its inventions when it filed a patent for more than 2,000 partial cDNA sequences last June. She noted that the two main associations representing virtually all the US biotechnology industry — the Industrial Biotechnology Association (IBA) and the Association of Biotechnology Companies (ABC) — support the NIH patent application. And she reproached other countries — specifically mentioning Britain — that have criticized the NIH application “but have not officially embraced a uniform policy of full disclosure or of patenting”.

The meeting was ostensibly to solicit public input for the deliberations of the committee, a genome patent working group that serves under the Federal Coordinating Council for Science, Engineering and Technology within the White House. But after nearly a year of debate, most groups with an

interest in the matter have already made their case, and the meeting, scheduled for two days, lasted only one.

Nevertheless, the working group intends to make sure that the government does not repeat its mistake of making a decision on this subject without getting input from every possible source. European and Japanese patent law gives the United States a 12-month grace period after a US patent application in which to file elsewhere. For NIH that deadline arrives on 20 June, and US officials promise that they will have adopted a coherent policy by then. Deciding not to file in Europe and Japan next month would be a policy reversal nearly as significant as the initial decision to file.

One entity, surprisingly, that would not mind seeing that kind of policy change is the US Department of Energy (DOE). David Galas, director of the DOE side of the US genome project and a member, with Healy, of the interagency patent committee, revealed tensions within the government on the issue.

“We’re not against patenting, or the patenting of genes”, he began. “Our reservations are otherwise.” He then listed several “unsettling aspects of the [NIH] patent”, including its “potential to inhibit collaboration and the free exchange of data, both nationally and internationally. We’re already observing this”, he noted, in apparent reference to the disputes between the NIH

and the British Medical Research Council (MRC) over access to MRC genome data (see *Nature* 354, 96; 1991).

Pursuing such patenting of partial gene sequences, Galas warned, could lead to further disputes and delays, could “inhibit real innovation” and could clutter the commercial front with patents that are “all but useless”. Another opponent of cDNA patenting who spoke at the meeting was Axel Kahn, research director of the French science agency, INSERM, and president of the French biomolecular engineering committee. Presenting the French government position, Kahn expressed “firm” opposition to patents on partial gene sequences, and accused NIH of exaggerating the risk of not patenting. NIH argues that allowing cDNA sequences to be published without intellectual property right protection robs industry of an incentive to develop products based on those sequences. That claim “seems to have no basis whatsoever in fact and, it must be said, was uttered to further the cause”, Kahn alleged. But few within US industry seem to be worried about that. In a letter sent last week to President George Bush, the ABC supported the NIH patent application and future sequences as “essentially the only responsible course under existing federal law” while the issue of patentability remains open. Following the IBA, which drafted a similar statement earlier this year, the ABC called for NIH to work with industry on a licensing policy for the sequences.

Given that a great deal of work is needed to turn a partial cDNA sequence with unknown function into a marketable product, the ABC asked NIH to license the sequences non-exclusively.

Christopher Anderson

British JET staff to strike over pay and prospects

London. Anger over unequal pay and fear about their future have led British workers at the Joint European Torus (JET) to the brink of a series of one-day strikes. The experimental nuclear fusion project, based at the Culham Laboratory in Oxfordshire, is undergoing an upgrade that is expected to be completed in 15 months, and union representatives at JET expect any job action to disrupt that timetable significantly.

Researchers from laboratories throughout Europe are hired to work at JET under the European Atomic Energy Community (EURATOM), and their salaries are set by a common scale. On the other hand, British scientists and technicians work at JET under the aegis of what was the UK Atomic Energy Authority (UKAEA), and are paid on a separate scale. In effect, a Euratom scientist earns twice as much as a British scientist for the same job.

As well as reflecting a long-standing dissatisfaction with this situation, the possi-

ble strike also signals growing uneasiness with employment prospects after the JET project ends in 1996. While European staff may return to the laboratories from which they came and apply from within EURATOM for the next phase of the fusion research programme, British staff must compete as external applicants. Traditionally, internal applications have the upper hand in bidding for jobs on such joint projects.

Officials from the union, the Institution of Professionals, Managers and Specialists (IPMS), say support for the strike is strong. Around 90 per cent of the 240 British scientists and technicians at JET — who make up half the workforce — belong to the union.

IPMS officials at JET have been campaigning for parity of pay for several years, but have made little headway. A petition to the European Parliament last year resulted in the Parliament directing the European Commission (EC) to set up an independent enquiry. The report is now overdue, and last

week a delegation from JET travelled to Brussels to try to move things forward.

A meeting with Fillipo Pandolfi, the EC’s commissioner for research and development, and Paolo Fasella, director general of the directorate that oversees such research projects, resulted in the promise of another committee, due to report in September. Union representatives felt that response was insufficient, and authorized a strike vote.

In addition to being unhappy about their prospects under EURATOM, British staff are also dissatisfied with UKAEA, which they were obliged to join in order to work on JET. UKAEA was funded entirely by the British government when JET was begun in 1978, but it is now responsible for generating much of its own income. Staff say that it no longer places such a high priority on research, and they are unhappy with the prospect of having to return to it once the JET project ends.

Ian Mundell

Canada's medical council wants larger role

Quebec. Next month, the Medical Research Council (MRC) of Canada is expected to endorse a long-term plan for a revolutionary change of role. From being a conservative, reactive fund-granting body little known outside biomedical research circles, it intends to become what its president, Henry Friesen, calls "the voice of Canadian medical and health-related research".

Beneath the lofty rhetoric, however, lies the down-to-earth question of whether the country can afford the council's new-found voice. And researchers worry that their funding will be squeezed unless the budget is increased.

The new direction would take the council far beyond the awarding of biomedical research project and research training funds after peer-review, its principal activity since its founding in 1969. It would make it the leading actor in all areas of health research, including epidemiological, disease prevention and environmental health studies as well as evaluation of treatment effectiveness. It would become, in effect, the Health Research Council of Canada.

Its decision grew out of a complicated consultative process involving more than 1,500 people. The issue was debated by some 100 representatives of the health sciences, government and the private sector at a national conference on 12-13 May, following 16 workshops at universities across Canada.

The expansion, which would be expensive, comes at a time when the federal government is trying to pay off a national debt of Can\$420,000 million (US\$300,000 million). Brian Mulroney, the prime minister, has identified science and technology as one of the building blocks of Canada's competi-

tiveness, but has found it difficult to show his support for science while keeping a tight rein on federal expenditures.

"The research component in our country makes us the finest country in the world", he told conference delegates. "I believe you and the institutes you represent, which look after our well-being, our health, our children and our scientific future, perform the most important work in Canada".

Unfortunately, Mulroney said, he inherited a situation upon becoming prime minister in which "we spend \$43 billion to service our debt on revenues of \$28 billion". The result, he said, is that "I don't have the resources to give you what you need".

Much was made at the conference of the fact that MRC and the other two research fund-granting councils have been promised an increase of 4 per cent a year for the next four years. What nobody mentioned was that the annual average inflation rate for the past year was 5.6 per cent.

Friesen thinks that, the gloomy financial picture notwithstanding, now is the time to enlarge MRC's scope because "health research must ultimately include the whole spectrum of health-related activities". Rigorous analysis of health care costs, he believes, could help to curb the steady growth of an annual bill, now \$70,000 million.

Friesen does not believe that the new plan will lead to a reduction in research funds for basic research, but rather the opposite. More partnerships and joint ventures with other government agencies, voluntary foundations and industry will lead to an increase in research funding sufficient to make the health industry a major contributor to the country's economy, he says.

David Spurgeon

India claims market drives US ban on space contracts

New Delhi. The Indian government says that a US ban on space contracts with the US government and companies is really intended to cripple its commercial space programme rather than uphold the principles of an international agreement to prevent the spread of nuclear weapons.

The United States last week imposed a two-year ban on exports to and contracts with the Indian Space Research Organization (ISRO) in retaliation for India's decision to continue its \$80-million contract with the Russians for cryogenic rocket engines. US officials believe that the contract violates the guidelines of the missile technology control regime (MTCR), which Russia has promised to follow, while the Indian government insists that the engines do not fall under the terms of the agreement (see *Nature* 357, 99; 1992).

The head of the Indian space agency, U.R. Rao, says that the United States did not object when a US company, General Dynamics, offered to sell the engines to India two years ago. ISRO rejected the offer because the cost was three times more than the bid by Glavkosmos, the Russian space agency, and the company did not promise to share the technology with Indian scientists. His agency has already paid one-third of the overall cost, and he said that the US pressure was "unjustified". The ban seems unlikely to harm India's rocket programme, Rao said. Three projects that rely very little on imported technology — an intermediate range ballistic missile and two satellite launch vehicles — are moving forward, as is the project to provide engines for the second stage of a geostationary launch vehicle.

At the same time, the ban could severely damage the country's plans to enter the commercial space market. Two remote sensing satellites now in orbit and a communications satellite to be launched next month by Ariane contain a significant amount of foreign technology. There is no way that India can quickly develop the ability to make these parts.

"The US embargo will no doubt mean a delay and cost overruns", says Rao, "but we will survive". Alternate suppliers may be scarce, he added, if other countries need a US call to impose their own ban on sales to ISRO.

If the embargo spreads, India may have to abandon its plans to become a global player in space. The ban on US contracts has already removed US tracking stations from their expected role in tracking next month's launch of the communications satellite INSAT-2A, as well as existing agreements to receive images from Landsat and NOAA satellites.

K.S. Jayaraman

Congress slashes "silly titles"

Washington. Defying scientific groups, who accused Congress of making a mockery of the peer review process, legislators last week voted to go ahead and seek the deletion of 32 projects with "silly titles" that had been merit reviewed and approved by the US National Science Foundation (NSF) and the National Institutes of Health (NIH) (see *Nature* 357, 103; 1992).

Although Congress left untouched many of the agricultural projects that President George Bush had initially marked for deletion, it — led by Senator Robert Byrd (Democrat, West Virginia) — gave no quarter to the NSF and NIH. All 31 of the NSF projects Byrd had targeted as "executive branch pork"

(including "Holism in psychobiology in the twentieth century" and an NIH study on why patients are afraid to go to the dentist) are slated to go. The language, however, may offer NSF a slight loophole — it requires a \$2 million cut but only "strongly suggests" that the money come from the 31 offending projects. NSF officials, who know on which side their bread is buttered, are wary of going against Congress' wishes. But they are reviewing their options to see if there is some way in which they can cut an equivalent amount elsewhere in their budget and leave most of the scorned grants untouched.

Christopher Anderson

CFCs active

SIR — Surely the description of CFCs as “greenhouse neutral”¹ is misleading. Although recent research suggests that these gases may cause only a slight change in the total amount of outgoing infrared radiation trapped by the Earth’s atmosphere², the heat will be trapped at different altitudes from those in the unperturbed atmosphere. This is not a neutral process, and if, for example, more heat is trapped in the lower atmosphere while the upper atmosphere cools it could stimulate enhanced convection and bring unpredictable changes in weather patterns.

John Gribbin

Rose Cottage, Piddinghoe,
Newhaven BN9 9AT, UK

1. Montgomery, K., *Nature*, **356**, 738 (1992).
2. Ramaswamy, V., Schwarzkopf, M. D., & Shine, K. P., *Nature*, **355**, 810–812 (1992).

Running in the family

SIR — We were interested to read the recent correspondence¹ about the sex ratio of artists’ offspring, as we have just finished a survey into the secondary sex ratio of male distance runners’ children, carried out because there is some anecdotal evidence to suggest that distance athletes father more daughters than sons.

Runners were approached at club training evenings and at races, in and around Glasgow. They were asked various questions regarding their physical activity around the time of conception of each of their children. The information sought included weekly training mileage, training intensity, whether heavily involved in other sports, and if they had any time off training due to injury.

We obtained data from 139 runners who were training around the time of the conception of at least one of their children, and from 66 males who took up running only after their children were born.

The results we found when the data were grouped by training mileage are summarized below:

Miles per week	Nil	0–30	30–50	50–70	70+
Sex ratio	0.62	0.63	0.40	0.58	0.53
Total no. of Children	181	80	116	60	36

Using the formula in ref. 1, the sex ratios of children born to fathers who were not running at the time of conception, and to those completing 0–30 and 30–50 miles per week were all signifi-

cantly different from the normal of 0.514 ($P < 0.001$, $P < 0.03$, $P < 0.006$ respectively).

None of the following could be related to sex ratio in any manner: training intensity, occupation (manual versus non-manual), paternal age, time between marriage and conception, and whether the father felt he was on a winning or losing streak when the baby was conceived.

Although we were relying on the athletes’ memories for their weekly mileages, the vast majority were very precise about their training regime at the time of conception. If there were some inaccuracies in the reporting of amount of training, it is difficult to see how these could appreciably affect the results.

The high sex ratio among the children born to fathers who later took up running could perhaps be due to their being more likely to become involved in the sport if their child was a son. Similar reasoning might be applied to those running more to keep fit rather than for competition. Thus they also may have been more likely to become involved if they had sons.

Many distance runners have lowered testosterone levels³, and the secondary sex ratio has been linked with parental testosterone levels at the time of conception⁴. Is this the reason for our result of a low sex ratio in middle distance runners? Do athletes covering distances in excess of 50 miles per week have higher than average testosterone levels and so father a greater proportion of sons? Further work in this area is desirable, not least to increase the numbers in our survey.

E. Crawford

D. Gilmore

W. H. James
University of Glasgow,
Institute of Physiology,
Glasgow G12 8QQ, UK

1. Beck, R. A. *Nature* **356**, 189 (1992).
2. Moore, D. H. & Gledhill, B. L. *Fertil. Steril.* **50**, 21–25 (1988).
3. Cumming, D. C. *Sports Med.* **7**, 1–17 (1989).
4. James, W. H. J. *theor. Biol.* **143**, 555–564 (1990).

SIR — R. A. Beck¹ proposes an observed sex ratio among the offspring of “artists” expressed as a small but significant male excess.

The sex-ratio is not a stable value, showing marked variation between populations as well as significant secular change. In the United Kingdom, it is currently the lowest recorded since 1933, with a steady decline since 1978 culminating in 1988 with a deficiency of males at 104.8 live births per 100 females, giving a sex-proportion of 0.5117 (ref. 2).

Converting Beck’s figures to the more usually adopted convention of sex proportion (the proportion of the total

population who are male) his value for “artists” is 0.5278 and for “others in *Who’s Who*” 0.5112.

There are many variables known to affect the sex ratios but one of the most significant is birth order³, with a declining sex ratio (fewer male births) as family size increases. Although slightly different from Novitski and Sandler’s US figures, this results in the United Kingdom in a sex proportion of 0.5294 for families of two children and 0.5271 for families of three children⁴, corresponding closely with Beck’s value of 0.5278 for artists.

If Beck looks at the family size as well as the years of birth, and thus birth order of his artists group, he may well discover family sizes of between two and three — possibly the maximum number that can be accommodated in an artist’s garret!

R. C. Connolly

Department of Human Anatomy
& Cell Biology, PO Box 147,
University of Liverpool L69 3BX, UK

1. Beck, R. A. *Nature* **356**, 189 (1992).
2. Shaw, C. *Popul. Stud.* **57**, 26–29 (1989).
3. Novitski, E. & Sandler, L. *Ann. Human Genet.* **21**, 123–31 (1956).
4. Connolly, R. C. (unpublished data).

British science

SIR — In their recent letter¹, Martin and Irvine merely rehash their old relative studies, and make no attempt to address the evidence of absolute British expansion in science. I shall therefore point out only that a relative British scientific decline is not only inevitable, it is actually desirable if it means that previously poor countries are catching up²; as long as British science grows absolutely at its current rates, the needs of the nation at large will have been more than met.

Let me correct two misstatements of Martin and Irvine¹. (1) My assertions have not been laid to rest by references 3–7 as detailed in my published rebuttal⁸. (2) I did not restrict my analysis of the Institute for Scientific Information’s most cited papers to 1983 and 1984 (see the legend to Table 3 in ref. 9), but the issues of *Current Contents* for 1990 and 1991 that cover citations in the late 1980s were published before my own paper was submitted to *Scientometrics*⁹.

Terence Kealey

Department of Clinical Biochemistry,
Addenbrooke’s Hospital,
Hills Road, Cambridge CB2 2QR, UK

1. Martin, B. R. & Irvine, J. *Nature* **355**, 760 (1992).
2. Kealey, T. *Science Technol. Policy* **4**, 8–14 (1991).
3. Roberts, D. H. *Nature* **350**, 550 (1991).
4. Ashall, F. & Goate, A. M. *Nature* **350**, 550 (1991).
5. Collins, P. *Nature* **351**, 8 (1991).
6. Humphreys, C. *Nature* **351**, 513 (1991).
7. Mulvey, J. *Nature* **351**, 513 (1991).
8. Kealey, T. *Nature* **352**, 466 (1991).

Italian university structure

SIR — Many readers not familiar with the structure of Italian universities may be confused by the recent controversy aired in these columns¹⁻⁵. As a foreigner (Australian) with several years experience of Italian academic life, I am perhaps well placed to try to explain some of the more unusual aspects of academic promotion that engender so much excitement and rancour in Italy.

Italians are renowned for their creativeness and adaptability, but their institutions are not, and the university system is sadly no exception. Despite some recent moves towards limited autonomy, all Italian universities remain firmly under the centralized control of the Ministry of Universities and Research. Individual universities can neither appoint nor promote their own professors. To promote a colleague, a department must apply to the ministry for a new position at the higher level, after lengthy and often heated debate by the full councils of the department, school of studies and faculty. If the request is granted, the chair is listed to be filled in the next *concorso*, the centralized selection procedure, with a total lead time of around five years. Vacancies from all universities are gazetted together in groups of similar disciplines (with no advertisements in *Nature* or the like). Applicants apply for whole groups of positions, not for individual chairs, expressing no preference for faculty, university city or region. This explains in part why many Italian academics, like the not atypical characters of David Lodge's *Small World*⁶, teach in cities remote from their abode: I myself live in Pisa.

For each group of positions, a commission of professors is elected by all professors in the relevant discipline. Of those elected for each commission, 60 per cent are eliminated *at random*, in a rare example of a selection procedure where randomness is considered to improve performance (by noise linearization?). Those remaining have the task of selecting the best applicants, considering only the scientific and didactic merits of the candidates. Full professors are selected on the basis of a written curriculum (without interview), while associate selection is a more gruelling affair, including a public lecture on an unfamiliar topic at 24 hours notice. Only after selection do applicants apply to the faculties of their preferred universities, who in turn select their preferred candidates (but in practice there is often no real choice⁵).

Now, does any reader believe that the system really works like this? Do the members of the departments and facul-

ties of the individual universities have so little say in the selection of their professors? Do candidates apply for pools of vacancies, with no preference for research institutions or prior knowledge of where they may end up? Of course not. Unless the preferred candidates completely disgrace themselves in the oral examination, the *concorso* is usually only a legal formality. Most of the real selection occurs behind the scenes, in huddled discussions at conferences and other clandestine encounters.

It is not hard to imagine how the discrepancy between the theoretical and actual selection criteria leaves the system wide open to criticism by disgruntled losers. Applicants who are best on paper may not be selected for a host of valid reasons, including local research, teaching and clinical requirements. However, if the real reasons are not legally acceptable, others must be invented, and this ambiguity can clearly lead to resentment. But the real problem arises from the fact that once one accepts that the law has to be bent to allow the system to work, it is a small step to bend it a bit further to make the system work for less honourable motives. As Alan Friedman⁷ convincingly argues for the business sector, academic institutions in Italy remain largely feudal. Many senior professors (still often referred to as "barons") view university chairs as fiefs to be awarded to loyal vassals (ensuring their eternal fealty) rather than to independently creative scientists. Competition for control of the fiefs is keen, with deals, vote-trading and other forms of favour-swapping, including grant distribution (for which there is no peer review). Alliances form along similar lines to those in the thirteenth century city states, and conspiracies and conspiracy theories abound.

I am in no position to express an opinion of the cases published in *Nature*¹⁻⁵, but there is no shortage of *concorso* anecdotes in Italy. For example, a close friend recently applied for promotion to associate-professorship. In the unofficial 'pre-*concorso*' hearings, her curriculum (including four first-author papers in *Nature*) led her to be judged a clear leader, and she so impressed the commission at her interview and public lecture that a commission member said that "she performed so well it will be extremely difficult for us to find a way to eliminate her". But they did, and all positions went to feal vassals of the commissioners and their allies. My friend was eliminated through no fault of her own, but because the commission considered that her *padrone* (who supervised her PhD 12 years ago but has not

since worked with her) had not participated fully in the *concorso* games.

It is difficult to assess quantitatively the degree to which promotion is more unfair in Italy than elsewhere. Certainly, we all believe that our own promotions are merited and, despite the unusual selection techniques, I suspect that the great majority are. However, the few blatantly outrageous promotions, combined with the ambiguity and hypocrisy inherent in the system, have created a general perception of corruption and nepotism, and this perception is seriously eroding morale. And the fact that sycophancy is more highly prized than originality is a real impediment to innovative research. *Concorso* power-mongering is a sport for the barons, with little consideration of its impact on the young (and often not-so-young) applicants who must survive the trial-by-fire oral examination, only to learn that the real selection was made months earlier on the basis of entirely different criteria. Why not dispense with the charade?

Most Italians agree that the system has major problems, and many talk of little else, but few agree on the remedy. Certainly it is unlikely that change will come from within. Not many professors seek the honour of serving on selection commissions (involving up to 20 days out of town, and maybe 150 hours of boring committee work) unless they have good reason to do so. But little is to be gained by vilifying the individual commissioners, most of whom are honest scientists trying to make an unworkable system work. My personal view is that the only way to restore some dignity to Italian academic promotion is to accept that centralized control is unrealistic, and to give some real autonomy to the universities, allowing them to select and promote staff to suit their local needs. Many cynics will argue that freedom from centralized control will lead only to increased nepotism, but I believe that this unfortunately widespread aspersion on the integrity of Italian scientists is unwarranted. The system is at fault, not the individuals. With appropriate mechanisms, such as linking university funding to performance, universities would soon see the advantages of choosing good people, and might even consider advertising for them.

David Burr

(Professore ordinario di Psicologia),
Università di Roma,
Via dei Marsi, 78, Rome, Italy

1. Gaetani G. F. & Ferraris, A. M. *Nature* **353**, 10 (1991).
2. Amadori, S. et al. *Nature* **355**, 58, (1992).
3. Baccarani, M. et al. *Nature* **356**, 188 (1992).
4. Auiti, F., Baroni, C. Cao, A. & Fantoni, A. *Nature* **356**, 188 (1992).
5. Visconti, G. *Nature* **356**, 188-189 (1992).
6. Lodge, D. *Small World* (Secker and Warburg, London, 1984).
7. Friedman, A. *Agnelli and the Network of Italian Power* (Mandarin, London, 1989).

Endangered species and the law

Valerius Geist

Taxonomy does not deserve its reputation as an arcane science. As the following examples from ungulate taxonomy show, classification has important implications for conservation legislation.

THE science of taxonomy may be thought of by many as an arcane subject, innocent of the problems of the real world. But when taxa became included in conservation laws, this innocence vanished. Endangered species acts, wildlife management laws and the Convention on International Trade in Endangered Species (CITES), essential, if imperfect, tools for conservation, label species and subspecies with formal taxonomic names. Such taxa, elevated to legal status, became subject to legal action. That may place biologists, accustomed to privacy, anonymous peer review and uncertainty in knowledge, in the witness stand, subject to hostile cross-examination. In one Canadian court case I was questioned for 24 hours on the taxonomy of red deer (*Cervus elaphus* Linnaeus, 1758). The case led to fines of \$25,000 for the defendant, who argued that by crossing two (invalid) subspecies listed in legislation he had created a new form of life, and that laws controlling commerce in wildlife did not apply to his creations.

Implications

It is not the image of scientists squirming in witness chairs that merits attention, rather the fact that courts and solicitors' offices are allowed to rule on taxonomy. Judges may now decide on matters such as the definition of species or subspecies, the criteria for establishing taxa, which taxa are valid, and which populations can be legally protected. The implications for conservation, but also for biology in general, are profound and

worrying.

Collectors, scientific or otherwise, are playing risky games taking chances on taxonomic uncertainties, for taxonomic mislabelling may now be a breach of law. That is no trivial matter if it leads to stiff fines or imprisonment, in particular if it brands the perpetrator a criminal. Even without such draconian consequences, the repercussions may be serious enough, as shown by the "Chinese argali" case (Fig. 1). Argalis (*Ovis ammon* Linnaeus, 1766) are giant sheep from central Asia, one of which, the Tibetan argali (*O. a. hodgsoni* Blyth, 1840), is on the endangered list of the US Endangered Species Act. Four Argali rams, shot in April 1988 by four US hunters in Ganzu Province, China for a payment of \$100,000, were confiscated by agents of the US Fish and Wildlife Service upon entry to San Francisco. The hunters had been accompanied by an employee of the service, who was on temporary loan to the Smithsonian Institution. The trophies were labelled merely as *O. ammon* on the export permit. The sheep were subsequently identified as Tibetan argalis by four experts, but this was disputed by the hunters.

Although the case was eventually settled out of court, it had wide repercussions. It led to hundreds of thousands of dollars in attorney fees, a Grand Jury investigation, a flurry of diplomatic activity between the United States and China, embarrassing publicity, serious rifts within the US Fish and Wildlife Service and within the International Un-

ion for the Conservation of Nature over policy, several international meetings on caprid conservation and formal taxonomic and status reviews. The case embroiled two secretaries of the interior, five senators, two congressmen, affected even Secretary of State James Baker, and led to frantic activity to 'de-fang' the Endangered Species Act by some hunters' organizations. Pressure was exerted on scientists acting as expert witnesses to change their testimony, and they were maligned or blackmailed. Nearly four years afterwards, the affair is not yet over.

Species identity

The use of alloenzymes and mitochondrial DNA analysis has thrown doubt on the identity of some species named in conservation legislation. Eastern timber wolves (*Canis lupus* Linnaeus, 1785) and the red wolf (*C. rufus* Audubon and Bachman, 1851) may carry the mitochondrial DNA of coyotes (*C. latrans* Say, 1823)^{1,2}, whereas Florida panthers (*Felis concolor coryi* Bangs, 1896) carry genes of pumas from central America because of a release of hybrids decades ago. These cases have generated concern since the US Solicitor's office of the Department of the Interior ruled that hybrids are not protected by the Endangered Species Act³⁻⁵. Although I agree with O'Brien and Mayr³ that the hybrid policy should be carefully applied on a case-by-case basis, and that hybridization need not be a tragedy for conservation, the unresolved question is whether the courts, rather than other

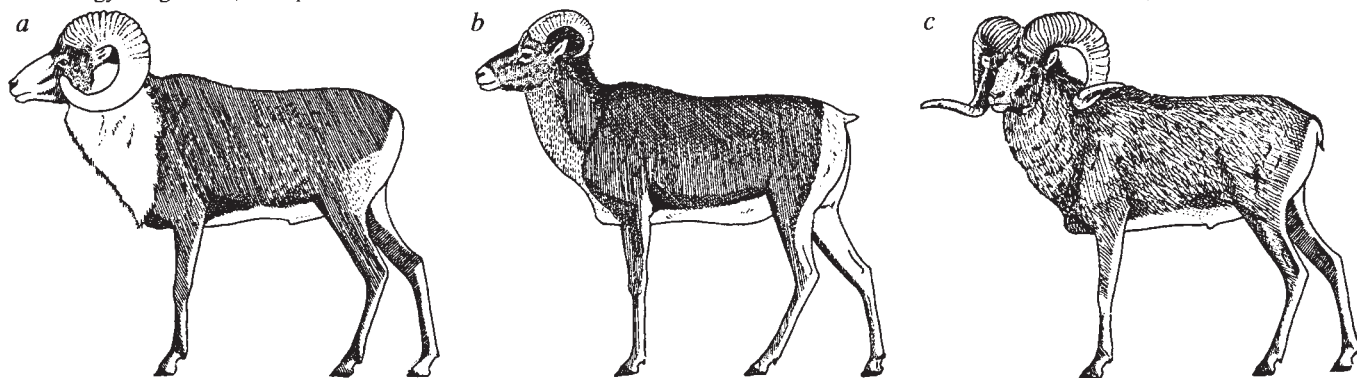


FIG 1 a, Mature Tibetan argali male (*Ovis ammon hodgsoni* Blyth, 1840) in nuptial pelage. There are at least 15 synonyms for this subspecies. b, Juvenile Tibetan argali male described as *O. a. dalai-lamae* Przewalski, 1888. The type specimen had the diagnostic dorsal neck ruff removed in mounting. It was later replaced with an altered mount of a mature *hodgsoni* male from another mountain range whose large horns had been removed and replaced with small horns. c, Mature Shansi argali (*O. a. jubata* Peters, 1876), negligently lumped with *O. a. darwini* Przewalski, 1884, due to a failure to read (but not to reference and denigrate) Peters' impeccable scholarship. The designation *jubata* has been applied to the *hodgsoni* specimen shot in 1988 in the Ganzu Province of China.

specialists, will concur with their view. Moreover, to succeed in court, there needs to be agreement on the criteria to use when applying the 'hybrid policy'. As I discuss below, one such criterion should be the success of hybrids as tested in their natural habitat.

Taxonomic flaws may have severe consequences for conservation, and Canada's 'wood bison' (*Bison bison* "athabasca" Rhoads, 1897) are a case in point (Fig. 2). These bison are enshrined as a formal subspecies in legislation, and have been subject to a longstanding, well-publicized national conservation effort. Recently, a consortium of agencies led by Agriculture Canada (including the interagency organization in charge of 'wood bison' conservation), proposed that the bison of Wood Buffalo National Park, as carriers of bovine brucellosis and tuberculosis, be eliminated and replaced with disease-free 'wood bison'. A federal review panel agreed with the consortium.

The proposal sparked controversy even beyond Canada's borders. The park bison are derived from about 1,500 native bison and 6,673 (diseased) southern-plains bison moved to the park in 1925. The 'wood bison' are captive offsprings from bison captured in the northern part of the park that were once thought to have escaped hybridization. If 'wood' and 'plains' bison are valid subspecies, then the park bison are hybrids and, as such, the consortium argues, are not worthy of conservation. Not only was the taxon *athabasca* based on an inadequate second-hand description of a single specimen and subsequently on worthless taxonomic methods, but when the animals were properly fed, the characteristic hair coat of 'wood bison' transforms into that of 'plains bison' within about a year.

Wood bison are, consequently, not a taxon, but an ecotype (and an artefact of captivity at that). Genetic analysis shows that 'wood bison' herds differ no more from plains bison herds than one from another. The designation of 'hybrid' for the park bison is thus false on taxonomic and genetic grounds. The species *B. bison* Linnaeus, 1758 has no subspecies, and that turns official conservation policy on its head. The park bison, still under selection pressure from predators, are not the worthless hybrids defined by officialdom, but are the most diverse, naturally tested gene pool remaining of the species *B. bison*. Replacing them with an inbred 'wood bison' ecotype, untested for decades by predators, would not only confuse phenotype with genotype, but would undermine the essence of conservation.

Clearly, taxonomy is important in a tangible sense. Yet worthless taxonomic methods, ignorance about the biology of

taxonomic criteria, failure to review original and foreign-language literature, faulty curation, tampering with museum specimens and labels, and artistic licence in illustrating taxa not only cast doubt on standards of scholarship in taxonomy⁷, but reduce the effectiveness of science in providing urgently needed legal protection to various forms of life. Already the American Society of Zoological Parks and Aquaria has expressed doubt about formal taxonomy, and is focusing instead on "genetically significant" populations as subjects of conservation. But this is no substitute, in my view, for an improved, formal taxonomy.

A fatal flaw in much large-mammal taxonomy is the use of comparative morphometrics as a taxonomic tool. Comparative morphometrics of crania or skeletons of free-living populations can no more be used to measure taxonomic (genetic) differences than a rubber band can be used to measure distance. Every set of comparable measurements conceals genetic, epistatic, environmental and statistical variation. That is, the gross variation is a mixture of different types of variation, within which the genetic variance is undefined. It remains indefinable, despite various approximations. Comparative morphometrics as a taxonomic tool is logically flawed. It confuses phenotype with genotype, analogy with homology, ecotype with taxon, and does not reveal the taxonomic and evolutionary differences between the populations compared. It reveals only differences, the origins of which remain obscure.

This flaw is not uncommon in other fields of biology⁸ when quantitative comparisons between populations are used to bolster evolutionary analysis. Such comparisons are futile if the proportion of variance attributable to heredity is unknowable. The closer the relationship between populations of a given form, the more speculative must be the conclusions about evolutionary relationships, because large phenotypic differences can arise from closely related genotypes in different environments. Taxonomic or evolutionary differences in close relatives should be studied experimentally, provided different variables affecting ontogeny are subject to effective control.

Taxa determination

In the absence of quantitative factors that are firmly (experimentally) based on genetic expression, what criteria might one use for determining taxa at the species and subspecies levels? The 'biological species concept' defines the species level in sexually reproducing organisms by using incompatibility of reproduction as an objective distinguishing criterion. Undoubtedly, this is a good and sufficient criterion in many

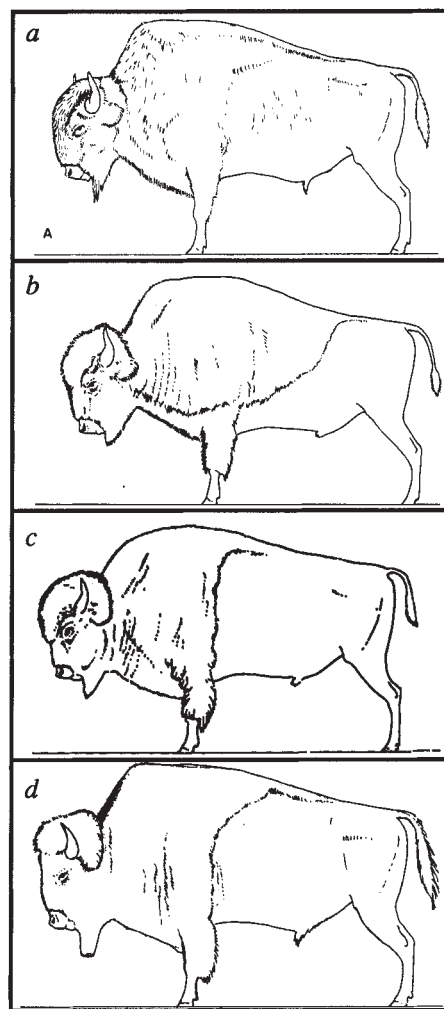


Fig 2 a, Classical 'wood bison' phenotype, Elk Island National Park. b, Bull from the same bison herd as a, but held and fed in a paddock in Banff National Park. c, Classical 'plains bison' phenotype. d, Wood bison bull from the same origin as a, but living free in the Mackenzie Bison Sanctuary. (From ref. 6.)

cases, particularly in species-rich communities with much sympatry. But this criterion, narrowly interpreted as the absence of hybridization in captivity by some⁹, is biologically flawed. This is painfully evident in the case of ruminants in the palaearctic and nearctic biogeographical domains, as widely differing forms hybridize readily in captivity, and the hybrids grow into adults under human care. But survival into adulthood is unlikely in natural conditions, where the hybrids are frequently exposed to predation. For instance, we found that hybrids of white-tailed (*Odocoileus virginianus* Boddaert, 1785) and mule deer (*O. hemionus* Rafinesque, 1817), which may develop in captivity into truly magnificent specimens, suffer from severe deficiencies when it comes to tactics and strategies of predator avoidance. The parent species have quite different means of escaping predators,

whereas hybrids have non-functioning mixtures of parental escape behaviours, plus inefficient locomotion. Escape strategies and tactics are, evidently, under close genetic control.

Is the current attempt to classify sika (*Cervus nippon* Temminck, 1837) and red deer into hybrid populations in several European localities^{10,11} an artefact of the absence of predators? In Manchuria, where both forms are sympatric, the populations remain distinct despite an occasional hybrid. Reproductive compatibility does not necessarily define a species, and sika and red deer differ in anti-predator strategies. Applying the conventional biological species concept would include in the same species hybrids that are not viable in nature. Viability of between-population hybrids under natural conditions of predation appears to be a better criterion for 'species'. It is a severe test, for even hybrids between subspecies may show loss of viability. This is suggested, for instance, by narrow hybrid zones, by ecological separation between some subspecies of Asiatic red deer, and by some difficulties in the hybridization of European red deer and North American wapiti.

Taxonomy without an understanding of ecology may have little relevance. Yet obtaining such understanding is difficult or even impossible today because of the widespread destruction of natural environments and the concomitant holding of large mammals in small populations in predator-free, artificial environments. The criterion of viability in natural environments will become increasingly difficult to test in the field, though experimental exposure of hybrids to predators, for instance, may be helpful. Consequently, species could be defined heuristically, as the next level of dissimilarity above the subspecies level, a practice adopted by those studying caprid biology¹².

Although O'Brien and Mayr³ defined subspecies geographically, this should be a fall-back criterion at best, as geographical origin has to be accepted on faith. For conservation laws to function effectively, one needs to be able to establish the geographical origin of specimens by factors intrinsic to the specimen. How could one spot that a protected species or subspecies was being imported without recourse to the label stating its origin?

The nuptial hair coat of reproductively active adults is an old, widely used taxonomic criterion for subspecies, but only in some lineages. Circumstantial evidence suggests that nuptial hair patterns are under close genetic control in almost all cases. Bison appear to be unique among large mammals, as the structure of their hair coat varies en-

vironmentally, just as do antlers in deer. The nuptial coat of adult mammals as a taxonomic criterion delineating subspecies can be applied accurately only with an understanding of sex, age and seasonal differences, as well as some knowledge of the effects of transplants to different environments.

Uniting populations with the same nuptial 'uniform' into one subspecies greatly reduces splintering of species based on meaningless morphometric differences. But it may obscure real genetic differences in ecological adaptations between populations which deserve to be recognized formally. This is difficult to accomplish, with few exceptions. One such exception is the bighorn sheep (*Ovis canadensis* Shaw, 1803) of North America. Although the 'uniform' is nearly the same in all sheep of this species, there are differences based on ecology, such as the short hair-coat, large ears and long tooth rows of sheep adapted to hot deserts. Some species, however, such as bears, cougar or white-tailed deer, are amenable neither to subspecific distinctions by nuptial dress, as they look much the same no matter where they live, nor to differences in morphology based on differences in ecological adaptations.

White-tailed deer, an ancient and very successful species, have a distribution from just short of the Arctic Circle in Canada, to 18° south of the Equator in Peru. Although they differ in such social features as presence or size of the metatarsal gland, or the relative size of the tail, or in ecological adaptations such as differences in the shedding patterns of the seasonal hair coats, these differences are overshadowed by the great uniformity in nuptial dress. Differences in size and antler morphology are not taxonomically meaningful, because these features vary greatly with environmental quality and seasonal factors. Yet, although remarkably uniform in external appearance, white-tailed deer in South America differ more genetically from white-tailed deer in North America than do white-tailed deer from black-tailed deer in North America. White-tails also differ genetically between regions, without these differences reflecting conventional 'taxonomic' differences^{13,14}. A white-tailed deer's geographical origin cannot be determined from its morphology, as can be done with reasonable accuracy for black-tailed deer, wild sheep, ibex or even red deer.

Species definition

How then does one define useful taxonomic units within such a species? Should one define subspecies genetically? That may be necessary, but it has pitfalls, as the Canadian bison controversy reveals. Genetic analysis shows that different

plains bison populations differ somewhat and are, roughly, equidistant from one another. All plains bison arose at the turn of the century from the same stock based on about 90 bison captured between 1873 and 1886. The differences between herds appear to be meaningless captivity effects, generated in part by the founder effect and subsequent drift in small herds. A further reduction in genetic diversity should arise in tiny founder herds when the dominant bull displaces others from breeding and, after breeding several cohorts of daughters, is displaced in turn by his larger sons. His sons go on to breed from their mother, aunts, sisters, daughters, cousins and so on. The male-dominance effect in small founder herds thus promises to narrow and distort genetic differences.

Should differences that arise from random factors and captivity effects be part of taxonomic nomenclature? One could generate 'subspecies' *ad infinitum* with every founding herd. We assume that the genetic differences between taxa are linked to adaptations to natural environments, otherwise taxonomy does not reflect the natural order of life. We must be careful to ensure that this order is not drowned out by the 'genetic noise' of man-made artefacts.

Clearly, the taxonomy of large mammals needs to be rethought. Taxonomy has gained greatly in importance since the advent of conservation legislation, and must be freed from neglect by proper funding and by intelligent practitioners. The dismantling of museum collections is not only tragic, but irresponsible, now that the specimens have gained legal significance. Without good collections and good taxonomists, signatories to international conservation conventions cannot adequately manage conservation within their own borders, let alone live up to their international obligations. □

Valerius Geist is in the Faculty of Environmental Design, University of Calgary, Calgary, Alberta, Canada T2N 1N4.

1. Lehman, N. et al. *Evolution* **45**, 104–119 (1991).
2. Wayne, R. K. & Jenks, S. M. *Nature* **351**, 565–568 (1991).
3. O'Brien, S. J. & Mayr, E. *Science* **251**, 1187–1188 (1991).
4. Gittleman, J. L. & Pimm, S. L. *Nature* **351**, 524–525 (1991).
5. Rennie, J. *Scient. Am.* **256** (4), 18–20 (1991).
6. Geist, V. *Arctic* **44**, 287–300 (1991).
7. Geist, V. *Can. J. Zool.* **69**, 706–723 (1991).
8. Hirsch, J. *Sage Race Relations Abstr.* **6**, 1–65 (1976).
9. Haltenorth, T. *Handbuch der Zoologie* **1**, (18), 1–67 (1963).
10. Bartos, L., Hyaneck, J. & Zivovnick, Zool. Anzeiger (Jena) **207**, 217–287 & 260–270; **208**, 30–36 (1981).
11. Harrington, S. *Inst. Forest* **30**, 64–78 (1982).
12. Nadler, C. F., Korobitsina, K. V., Hoffmann, R. S. & Vorontsov, N. N. *Z. Saugetierkunde* **38**, 109–125 (1973).
13. Smith, M. H., Baccus, R., Hillstead, H. O. & Manlove, M. N. in *White-tailed Deer* (ed. Hails, L.) 119–128 (Stackpole, Harrisburg, 1985).
14. Smith, M. H., Brannan, W., Marchinton, R., Johns P. & Wooten, M. J. *Mammal* **67**, 103–111 (1986).

'alala live on in Hawaiian forest

Contention over the fate of a dozen endangered crows in Hawaii seems to have been dispelled (for the time being) by sweep reason and a US National Academy of Sciences inquiry.

In the koa forests of the McCandless ranch on the Big Island of Hawaii live eleven 'alala, the last known members of their kind in the wild. Another ten 'alala live in a breeding colony on the island, but their success at producing chicks has not been great. Once a dominant species of crow (*Corvus hawaiiensis*) in the Hawaiian archipelago, conservationists call the 'alala a "classic indicator species" or model for efforts to save as many as twelve other species of Hawaiian forest birds from extinction.

As all too often these days, the preservation of endangered species is a challenge not only to science but also to the legal system as proponents of one course or another end up in court. And so it has been with the 'alala, whose case has been a *cause celebre* in Hawaii for the past decade. In an unusual but gratifying turn of events, the long-running controversy appears to be on the verge of resolution thanks to an independent scientific analysis of the issues by a committee of the US National Academy of Sciences (NAS).

It seems that various plans to save the 'alala in the wild, drawn up by both federal and state officials, were never implemented because the owners of the 64,000 acre McCandless cattle ranch, which is home to 11 'alala, refused to allow authorities on the property. The ranchers, acting on the advice of their own scientific advisers, took the position that the best way to preserve the 'alala was to leave them alone. Environmentalists, on the other hand, wanted to capture the birds and send them to the breeding colony on the theory that the genetic diversity of the wild 'alala would be just what was needed to improve reproductive success of the captive colony.

It was at this juncture that the US National Audubon Society and its Hawaiian affiliate went to court, suing the US Fish and Wildlife Service for its failure to protect the 'alala under the Endangered Species Act. Audubon argued that fish and Wildlife ought to have the authority to overrule the private property interests of the McCandless ranch owners. At this point, Fish and Wildlife persuaded the NAS to step in with a thorough analysis of the literature and a site visit to the ranch. The outcome has been something of a surprise.

The NAS committee, chaired by W. Donald Duckworth of the Bishop Museum in Honolulu, took issue with the logical presumption that there is a connection between reproductive failure in the breeding colony and the inevitable loss of genetic

variation that accompanied the decline of the once thriving 'alala population.

"Recent DNA analysis shows the captive flock to be moderately to highly inbred," the Duckworth committee concluded, but added that because the wild birds on the McCandless ranch live in close geographical proximity to the forests from which the captive birds were taken "the addition of wild adult birds from the ranch...for genetic

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

An adult 'alala glistens in the sunlight.

reasons alone would be expected to provide no more than a very minor measure of new genetic variation...."

The committee's analysis also contradicted assumptions that the McCandless birds were dying off at an unusual rate. Recent sitings of banded birds on the ranch revealed that the survival of adult birds is consistent with that for other crows. (By contrast, adult 'alala off the ranch had become extinct during the past two decades.) The data showed the reproductive rate of the McCandless birds to be normal, given the small size of the population, and attributed reproductive failure during the 1970s and 1980s among other wild populations to adult mortality rather than breeding incapacity.

"These observations amounted to an entirely fresh interpretation of the data", says Donna M. Gerardi of the NAS Board on Biology that managed the study.

The committee's analysis of the broader issue of species decline and habitat loss also produced some fresh observations. It is, for instance, logical to associate a loss of forest habitat and, in this case, the fruit-bearing trees that support the 'alala with the bird's imminent extinction. The forests of the McCandless ranch, where 'alala nest in the upper branches of the 'ohi'as trees and feed on fruits such as the 'ie'ie, mamaki, ho'awa and pilo, offer a natural habitat for the

birds. "A dramatic correlation between loss of understory habitat and decline of an 'alala population occurred in the late 1970's on Hualalai [island]," when understory vegetation was removed and exotic pasture grasses were planted, the committee observed.

However, its analysis of the history of the 'alala also revealed that the bird's population decline began before 1900 in areas that were not cleared for ranching, thereby ruling out a simple cause and effect relationship between habitat alteration and the 'alala.

The NAS's review of other environmental causes of the birds decline — the introduction in Hawaii of the predatory mongoose, for instance, or the incidence of viral or bacterial disease — also ruled out clear-cut explanations for the accumulating loss of the 'alala during the past 90 years. Wisely, the committee concluded that it could not explain the decline and said: "Even if a great deal is known about a species, we might not know the right things."

In the end, the Duckworth committee posed a compromise resolution that seems to commanded the respect of all parties. Essentially supporting the McCandless owners' view that the 'alala should be left alone, the NAS nevertheless recommended that researchers be permitted to take the first clutch of eggs from the nest every spring for hatching in the breeding colony. Only outstandingly temperamental females fail to produce a second clutch after losing their first, so that the population both in captivity and in the wild will thereby be increased. "Ranchers often know and understand the natural history of their properties extremely well," the committee concluded, and "need to be respected" in the process of protecting endangered species. (Throughout much of the dispute the ranch owners were cast as the bad guys for keeping the good-guy environmentalists off their property.)

Beyond the recommendation to retrieve the first clutch of eggs, there are the expected recommendations for more and better facilities for studying the 'alala and for continued analysis of its habitat.

Meanwhile, the parties to the lawsuit are busy studying the NAS's study and, according to all reports, are approaching common ground. The lasting lesson of the case is that dispassionate scientific analysis has challenged the prevailing wisdom about the role of genetic variation in preserving the species and may just have brought about an unexpected compromise that everyone involved (the 'alala included) can live with.

Barbara J. Culliton

Paul J. Conry

Bottoms up for the oceans

Robert M. May

THE comparative neglect of biological diversity on the oceans' floor finds redress (or at least initial redress) in analyses by Grassle and Maciolek published in *American Naturalist*¹. They say that the marine 'macrofauna' — mostly molluscs, crustaceans and polychaete worms — may number 10 million species, and that "this estimate is probably conservative". But whether or not their figures are broadly correct perhaps matters less than a message implicit in their study.

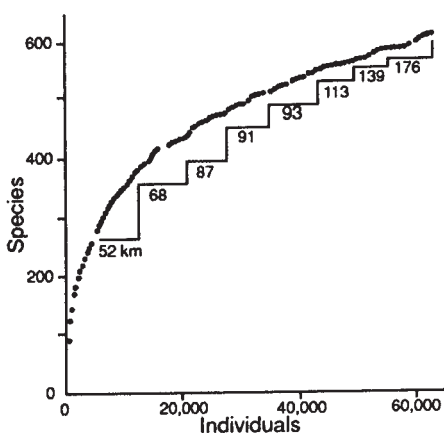
Most of the recent estimates of the total number of species likely to exist on Earth today, from Erwin's² widely cited 30 million to more cautious figures of 3–8 million^{3–6}, deal almost exclusively with terrestrial plants and animals. Terrestrial insects (which constitute more than half of all named and recorded species) have been the centre of attention, although other groups, such as fungi^{7,8}, nematodes and bacteria^{9,10}, have their fans. The marine realm has been largely ignored, for two reasons. First, it is estimated that of all the species currently named and recorded, fewer than 15% are to be found in the oceans¹¹, most of them in or on ocean-floor sediments. Second, most taxonomists and ecologists have been inclined to accept the recorded number of marine species as broadly representative of the true total, and not to expect the kinds of dramatic upward revision proposed for insects or fungi. This expectation, in turn, is partly because samples taken from poorly studied areas have tended to contain high proportions (typically more

than half) of species known already^{11,12}, and partly because the oceans are often seen as "a vast desert, desperately short of nutrients and with living things spread most thinly through them"¹³.

Enter Grassle and Maciolek, who base their figures on sampling of deep-sea communities on the continental slope and rise off the east coast of the United States. They took 233 'box core' samples (each covering an area 30 cm by 30 cm) from the ocean floor at depths of 1,500 m to 2,500 m along a 176-km transect off the New Jersey and Delaware coastlines. There has been no previous study on this scale of the invertebrate fauna living on or in the sediments of the deep-ocean floor. From a total of 90,677 individuals, Grassle and Maciolek identified 798 macrofaunal species, representing 171 families and 14 phyla; an additional 64 species were excluded because they were planktonic or lived on hard surfaces, not soft sediment. Of the 798 species, 46% were annelids (segmented worms, mainly polychaetes), 23% were arthropods (mainly various kinds of crustacean), and 13% were molluscs. At the family level, each of these three groups accounted for about a quarter of the total.

The 233 samples come from a total of 14 stations, mainly at depths of 2,100 m, sampled at various intervals over a two-year period. The relative abundances of the species from the nine 2,100-m stations sampled over the entire two-year period are remarkably similar from station to station and from time to time. Moreover, about 20% of the species were found at all ten 2,100-m stations, whereas only 28% of the entire soft-sediment fauna occur once and 11% twice.

Using these analyses as a springboard, Grassle and Maciolek combined samples over time at each site to examine patterns in adding new species as they moved in a southwesterly direction along their 176-km transect. They find a fairly typical 'rarefaction curve' in the numbers of species added (see figure) as they accumulate miles (or total numbers of individuals) along the transect. As always, the rate of adding new species is rapid at first (the first 500–70 km or the first 20,000 individuals in the samples), but then slows down. Grassle and Maciolek suggest that, in this later pattern, new species are being added to the total at around 100 species per 100 km. This, remember, is along a contour of roughly constant depth. Such evidence as they and others¹⁴ have compiled implies that rates of adding species across depth contours will be greater than the rough 1 species for 1 km along contours.



'Rarefaction curve', obtained by combining samples (over time within stations, and then over stations along Grassle and Maciolek's northeast-southwest transect). The distances mark transitions from station to station along the transect. The curve shows the rate at which new species turn up as more and more individuals are sampled. (From ref. 1.)

It is on this basis that the authors venture their very tentative marine total: if a linear rate of one new species per kilometre is generalized to one new species per square kilometre, then the oceanic area of 3×10^8 km² deeper than 1,000 m would mean that there are a few hundred million species of bottom-dwelling (benthic) marine macrofauna. But recognizing that the deepest and most oligotrophic parts of the ocean have densities of living things about an order-of-magnitude lower than found in their continental shelf samples, they scale their estimate of the total number of benthic species back to 10^7 or so.

Grassle and Maciolek conclude by pointing to ecological reasons for oceanic sediments harbouring a fauna as diverse as this. First, the input of nutrients to deep-sea sediments is inherently a patchy and ephemeral process; terrestrial ecologists have emphasized that such processes can be important promoters of species diversity^{15,16}. Second, sediment dwellers create small-scale disturbances which can increase environmental heterogeneity, and thence diversity. And all this is, they believe, compounded by the lack of barriers (compared with terrestrial environments) to long-distance dispersal, which allows distant migrants to contribute to reshuffling patterns of local diversity.

I have two worries about Grassle and Maciolek's estimate. If ease of immigration and long-distance dispersal are important cogs in the machinery maintaining overall diversity, then you cannot extrapolate a 'local' rate of adding species (for example 1 species per kilometre) beyond the characteristic distance scale on which the dispersal/diversity mechanism operates. This undercuts the extrapolation from the New Jersey/Delaware coast to the entire ocean. Quite apart from that, I am unhappy about calling the later part of Grassle and Maciolek's rarefaction curve a straight line for the purpose of making

1. Grassle, J. F. & Maciolek, N. *J. Am. Nat.* **139**, 313–341 (1992).
2. Erwin, T. L. *Coleopt. Bull.* **36**, 74–75 (1982).
3. Gaston, K. J. *Conserv. Biol.* **5**, 283–296 (1991).
4. Stork, N. E. *Biol. J. Linn. Soc.* **35**, 321–337 (1988).
5. May, R. M. *Phil. Trans. R. Soc. B* **330**, 293–304 (1990).
6. Hodkinson, I. D. & Casson, D. *Biol. J. Linn. Soc.* **43**, 101–109 (1991).
7. Hawkesworth, D. L. *Mycol. Res.* **95**, 641–655 (1991).
8. May, R. M. *Nature* **352**, 475–476 (1991).
9. Ward, D. M., Weller, R. & Bateson, M. M. *Nature* **345**, 63–65 (1990).
10. Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. & Field, K. G. *Nature* **345**, 60–63 (1990).
11. Barnes, R. D. *Am. Zool.* **29**, 1075–1084 (1989).
12. Hammond, P. M. in *Global Biodiversity* (ed. Groombridge, B.) 19–37 (Chapman & Hall, London, 1992).
13. Colinaux, P. *Why Big Fierce Animals are Rare* (Princeton University Press, 1980).
14. Blake, J. A. et al. *Biological Processes on the US South Atlantic Slope and Rise* (US Dept Interior, Washington, DC, 1987).
15. Atkinson, W. D. & Shorrocks, B. *J. anim. Ecol.* **50**, 461–471 (1981).
16. Ives, A. R. *Ann. Zool. Fennici* **25**, 75–88 (1988).

a global projection — with the same argument, a road atlas can be used to prove that the world is flat.

A more serious objection comes from looking at the fraction of species new to science in the samples described in the article. Of the 798 total, 460 species (58%) are not previously recorded. Within the most abundant group, the polychaete worms, 64% had not been described before; the molluscs were better known, with 37% undescribed. Taking these figures as representative, I would be inclined to make the more direct guess that the total number of species of benthic macrofauna is about twice that currently recorded, and unlikely to exceed half-a-million. The counter-argument could reasonably be made that the deep-sea fauna off this part of the US coast is better known than most, and therefore an estimate made in this way is low.

Whether the overall total of species of worms, arthropods and molluscs in deep-sea sediments is more than 10 million or less than half-a-million, Grassle and Maciolek's work is a reminder of how little is known about the two-thirds of the Earth's surface that lies below the deep oceans. Even if this region does indeed contain only 15% of the species on our planet, it arguably contains more information about the evolution of life on Earth than does the land: at the fundamental level of diversity of basic body plan, more than 80% of all phyla are found only in the sea, most of them in sediments. That alone, the message implicit in the study, is good reason for more taxonomists to turn their attention to the oceans. □

Robert M. May is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

PEPTIDE RECEPTORS

Stepping up the pressure

Mohammed Sharif and Michael R. Hanley

THE molecular cloning of V2 vasopressin receptors has been doggedly pursued for several years. As reported by two groups in this issue, the endeavour has now met with success — Birnbaumer *et al.*¹ (page 333) describe the predicted sequence of the human V2 vasopressin receptor, Lolait *et al.*² (page 336) that of its counterpart in the rat.

Vasopressin, one of the most intensively studied of peptide hormones, has a broad range of effects on mammalian tissues and may act at several G protein-coupled receptors. But its one undisputed physiological function is as a circ-

ulating hormone that regulates water retention by the kidney (hence the alternative name 'antidiuretic hormone'). Renal action seems to be mediated through a specific receptor, the V2 vasopressin receptor, which is thought to be restricted to the kidney and which — intriguingly — co-localizes at the chromosomal locus for nephrogenic diabetes insipidus^{2,3}. This is a rare, X-linked recessive genetic disorder, exhibiting renal insensitivity to vasopressin. So here we may have an instance of failure of a G protein-coupled hormone receptor being implicated in an inheritable disease.

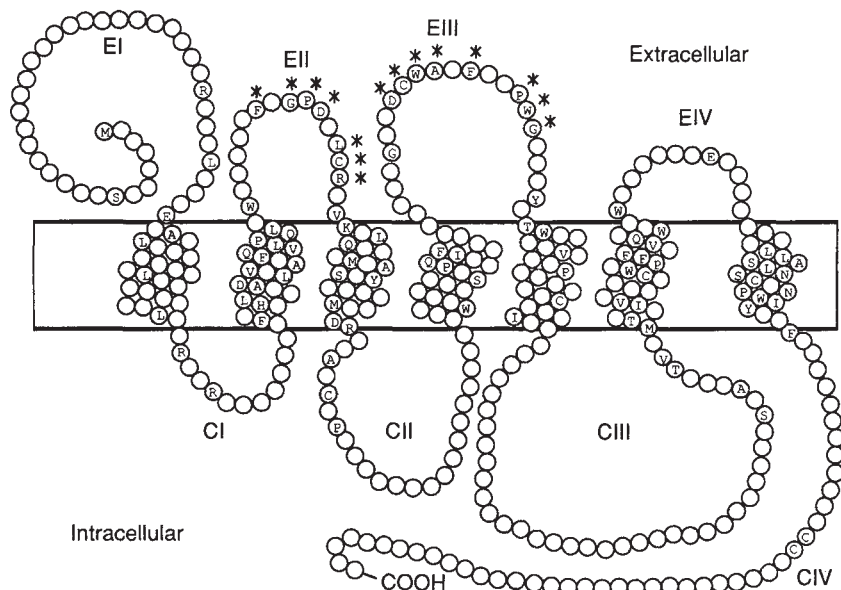
The interest of the two papers goes both further and broader, however, in that it turns out that V2 vasopressin receptors have considerable sequence similarity with the V1 vasopressin⁴ and oxytocin receptors⁵ reported last month. This similarity, although widely anticipated, was not a foregone conclusion, and with it comes practical implications for the analysis of receptor structure and function.

The significance of similarities between these receptors is that they are mirrored by similarities in the peptide hormones themselves. Arginine vasopressin and oxytocin are surprisingly alike (see figure), sharing seven out of nine residues and having an essential disulphide-linked ring structure which constrains the flexibility of the peptides⁶. Nevertheless, the peptides have distinct biological actions (for instance, through the V1 receptors vasopressin contracts vascular smooth muscle, whereas oxytocin contracts uterine smooth muscle) which arise from genetically distinct receptors.

So how do very similar, and conformationally constrained, peptides distinguish between two closely related receptors? Accordingly, can alignment of the cloned sequences give any insight into a possible shared recognition domain? Two identical regions of predicted extracellular sequence, not shared with other receptor subfamilies, are found in the EII and EIII domains (asterisks in figure). These conserved regions could define a spatially contiguous domain, stabilized by a conserved predicted disulphide bond, and are a plausible candidate for part of the peptide-recognition site. Although it is still unknown how peptides as a class may interact with their cognate receptors, this region of sequence similarity points to the EII and

- a Cys-Tyr-Phe-Gln-Asn-Cys-Pro-Arg-Gly-NH₂
 b Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly-NH₂
 c Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Arg-Gly-NH₂

Above, sequences of the (a) arginine vasopressin, (b) oxytocin and (c) chimeric vasotocin peptide hormones. Right, the predicted domain structure of the human V2 vasopressin receptor, identifying residues which are identical within the subfamily. Conserved residues in the extracellular EII and EIII domains are marked by asterisks. The predicted extracellular domains are identified as EI to EIV, where EI is the amino-terminal sequence, and predicted cytoplasmic domains are identified as CI to CIV, where CIV is the carboxy-terminal sequence.



EIII regions as targets for mutational analysis.

The relationship of species-specific hormone variants of vasopressin and oxytocin, such as lysine vasopressin in pigs, to possible variation in receptor sequences provides an excellent opportunity for matching naturally occurring peptide structures to variant receptor structures, and building trees of receptor evolution. In this regard, the transcript size for the V2 receptor from the pig-derived LLC-PK1 cell line is smaller than that for rats or rabbits¹. A porcine 'lysine vasopressin' receptor might have precise differences in its primary sequence relevant to speculations about where peptides might bind.

The question of a species-specific variant lysine vasopressin receptor touches on another issue. How many more members are to be found in this receptor subfamily? A 'V1b' receptor, the pituitary V1 isotype, is one obvious cloning target, but hybridization of a V1a receptor probe to Southern blots of rat DNA gave five to nine bands⁴, implying that the subfamily is much bigger than was previously suspected. For example, are there mammalian receptors selectively responsive to the chimaeric hormone 'vasotocin' which is found widely in non-mammalian vertebrates? Certainly, as the receptors proliferate, perhaps the problem of new peptide hormones should also be revisited; there is, for instance, evidence for an as yet uncharacterized neuronal vasopressin-like peptide⁷. Ultimately, are the multiple vasopressin receptors physiologically responding to vasopressin, or — as was found for tachykinin receptors — are the multiple receptors matched to multiple hormones⁸?

There are comparatively few examples of hormone receptors confined to a single organ, but the V2 vasopressin receptor is often regarded as unique to the kidney. Genetic probes now offer the possibility of obtaining conclusive evidence for or against its extrarenal expression. Unfortunately, the choice of tissues shown for northern blot analysis of V2 receptor messenger RNA expression² leaves this question open. The best cases for extrarenal expression of the V2 receptor have been based on haemodynamic or coagulant effects, not seen in nephrogenic diabetes insipidus,

of the selective V2 receptor agonist, 1-desamino-8-D-arginine vasopressin, on human patients⁹.

Molecular cloning has revealed that vastly different biological messengers, from lipid mediators to odorants to photons, trigger receptors based on an identical domain design consisting of single polypeptide chains predicted to have seven transmembrane α -helices. At the level of primary sequence, the number and diversity of these receptors is simply overwhelming. The challenge is to organize this information into some sort of order. Sequence alignment may be restricted to assessing evolutionary

relationships. It is not yet clear whether functional similarity, such as sharing a common ligand or a common transduction pathway, can be used as a predictor of specific structural relationships. So the molecular characterization of the members of the vasopressin/oxytocin receptor family promises to open an exciting new frontier for the Magnificent Seven. □

Mohammed Sharif and Michael R. Hanley are in the Department of Biological Chemistry, School of Medicine, University of California Davis, Davis, California 95616-8635, USA.

ATOMIC PHYSICS

Clarity in ion traps

Daniel Segal and Richard Thompson

IONS suspended in the field of a low-energy radiofrequency ion storage ring can be made to order themselves into unusual linear and spiral crystalline arrangements, H. Walther and colleagues report on page 310 of this issue¹. The results will no doubt be of interest to high-energy physicists who are keen to find ways of packing more ions into their accelerators and of compressing the ions' velocity distribution. But it is those working on high-resolution spectroscopy, particularly with a view to improving atomic clocks, who will benefit first.

The storage ring used in this work is a close relative of the radiofrequency (r.f.) quadrupole ion trap; work by W. Paul and H. Dehmelt using this type of trap was rewarded with a share in the 1989 Nobel prize for physics. The theme of the prize that year was high-precision atomic spectroscopy and particularly its application to the realization of frequency standards or atomic clocks. In an r.f. quadrupole trap a small cloud of ions, or even a single ion, can be confined to a small, spheroidal volume at the centre of an axially symmetric set of electrodes, to which an r.f. potential of several hundred volts is applied. It is the technique

of laser cooling² which has really opened the door to ultrahigh resolution spectroscopy in the trap environment; the cooling minimizes the ions' motion to yield essentially 'Doppler free' conditions.

Indeed, a single Hg^+ ion confined to the centre of an r.f. trap has been cooled strongly enough that it spent 95 per cent of its time in the quantum-mechanical ground state of translational motion in the effective potential well generated by the trap³. When, as was the case in that experiment, an atomic particle is confined to a region of space much smaller than one wavelength of the light used to observe it, it is said to be in the Lamb-Dicke regime. In this regime, the first-order Doppler broadening of the atom's spectrum is eliminated. The concomitant near-elimination of the second-order (relativistic) Doppler effect, combined with the long interrogation time possible in the trap, points to exciting prospects for ultrahigh-precision spectroscopy and optical frequency standards.

The limitations of such a frequency standard would be associated with the relatively low signal level that can be obtained in precision experiments on a single atomic particle. Walther's group

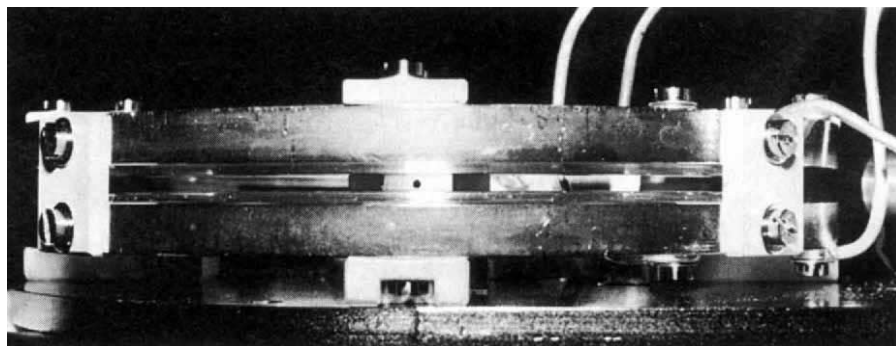


FIG. 1 The circular ion storage ring used by Walther and colleagues, with the atomic-beam inlet off to the right. The whole structure is 11.5 cm across.

1. Birnbaumer, M. *et al.* *Nature* **357**, 333–335 (1992).
2. Lolait, S. J. *et al.* *Nature* **357**, 336–339 (1992).
3. Jans, D. A., Van Oost, B. A., Ropers, H. H. & Fahrenholz, F. *J. biol. Chem.* **265**, 15379–15382 (1990).
4. Morel, A., O'Carroll, A.-M., Brownstein, M. J. & Lolait, S. J. *Nature* **356**, 523–526 (1992).
5. Kimura, T., Tanizawa, O., Mori, K., Brownstein, M. J. & Okayama, H. *Nature* **356**, 526–529 (1992).
6. Hruby, V. J. in *Topics in Molecular Pharmacology* (eds Burgen, A. S. V. & Roberts, G. C. K.) 99–126 (Elsevier-North Holland, Amsterdam, 1981).
7. Hanley, M. R. *et al.* *Nature* **309**, 258–261 (1984).
8. Maggio, J. A. *Rev. Neurosci.* **11**, 13–28 (1988).
9. Bichet, D. *et al.* *Kidney Int.* **36**, 859–866 (1989).

at Garching was among the first to show^{4,5} that small clouds of only a few ions can be made to undergo a phase transition from a chaotic cloud state to an ordered structure or 'crystal', by the application of strong laser cooling. This begs the question; why not use a cooled ion crystal in a trap, instead of a single ion, to create the clock?

The problem is that only an ion confined to the very centre of the quadrupole trap can be cooled sufficiently. This is because of the micromotion, which is a residual motion driven by the applied r.f. trapping voltage. The micromotion disappears at the centre of the trap but increases with increasing distance from the centre. When there is more than one ion in the trap, the Coulomb repulsion ensures that they take up lattice positions away from the trap centre, laying them open to the micromotion.

The ring trap described in this issue (Fig. 1) may provide one response to the challenge of cooling a large number of ions down to the Lamb-Dicke regime: what was a unique point in the three-dimensional quadrupole trap, the trap centre, becomes a circular line in the altered geometry of the storage ring, which is in effect a two-dimensional trap. In such a ring a very large number of ions (several thousand) can be stored in a string along the centre line of the electrode structure. All of these ions are now in a position where they are not subject to micromotion.

Sufficiently cold strings of ions hold great promise for a number of different experiments. Walther *et al.* recently suggested⁶ that such a string would be subject to the Mössbauer effect: the recoil due to the photon's momentum would be taken up by the string as a whole, not by the individual emitter, so avoiding the associated reduction in spectral resolution. Novel laser cooling schemes that depend on the multi-level structure of the cooling transition have been suggested for use in the quadrupole trap environment⁷. These new schemes hold the promise of being able to cool ions down to temperatures at which such effects could be observed.

The structures described in this issue were observed by focusing the fluorescent light from the ions onto an imaging photon detector system. The type of structures observed when the Mg^+ ions are cooled depends on a single parameter, λ , which is the normalized linear density of ions around the ring. If λ is low enough, a simple string of ions arises, which is the structure most interesting to atomic physicists. As λ is increased, either by increasing the number of ions or by reducing the strength of the transverse confinement, a point is reached where the string buckles into a zigzag structure. For the trap used by

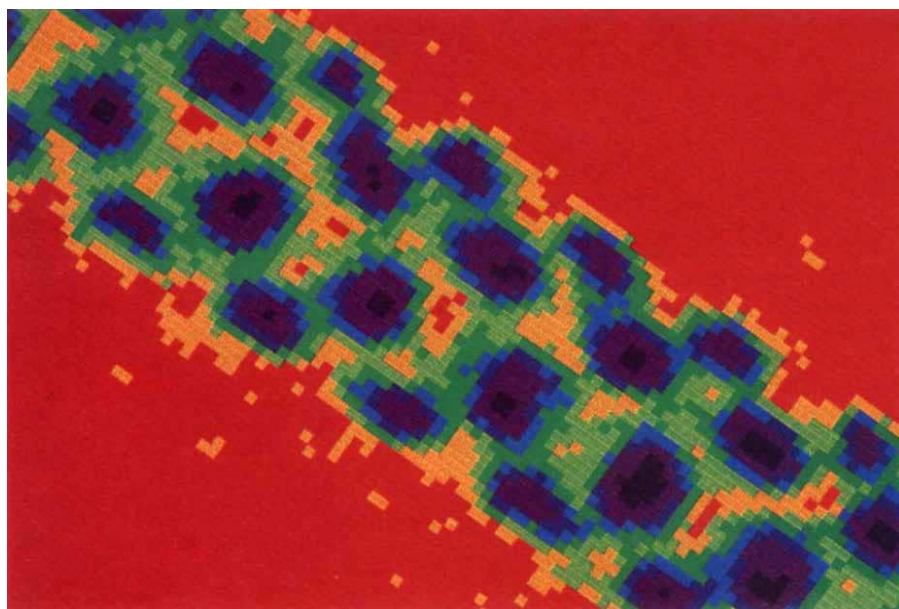


FIG. 2 Triple-helical ion crystal caught in the Garching trap. Its diameter is 53 μm , and the crystal is 15 orders of magnitude more rarified than its better-known solid-state cousins.

Walther and colleagues, this occurs when the number of ions in the ring exceeds roughly 10^4 . Increasing λ further gives more transitions, first to interwoven helices and finally to multiple shells (Fig. 2). Individual ions can be resolved only for the simpler structures, but the overall pattern can be seen for up to four shells. Theorists working on this problem⁸ will be pleased to see that their detailed predictions are borne out in these experiments.

It is fortuitous that the ions can be imaged at all. At the inlet for the atomic beam, used to load ions into the ring, a deposit of magnesium has built up on the gold plated electrodes which causes the appearance of a contact potential. This 1-volt potential acts as a lock on the ion crystals: without it the ordered structures would be free to rotate around the ring, smearing out any images one might try to obtain. The dependence on this potential barrier highlights the link between this work and that of a group at the National Institute of Standards and Technology (NIST) in Boulder, Colorado. Those workers have observed similar ordered structures of $^{199}Hg^+$ ions in a linear quadrupole trap, which looks a bit like a short section of the Garching storage ring. To stop ions from escaping at the open ends of the trap, extra electrodes are added at either end, to which a small positive potential is applied. These extra electrodes then essentially perform the same function as the contact potential did in the storage ring. In this linear trap, ordered structures with only a few tens of ions have been observed.

The NIST workers stress the importance of their trap in the field of high-

resolution spectroscopy, pointing out that imaging techniques could allow each ion to act as an independent atomic clock. Each clock could be observed with essentially 100 per cent efficiency by monitoring its 'quantum jumps'. They have used their trap for a pilot experiment⁹ in which they used the Ramsey 'separated-field' method in the time domain to measure the frequency of a microwave transition to about one part in 10^{12} . In the ultrahigh-precision world of atomic clocks, perturbations from a number of surprising sources can actually be greater than the residual second-order Doppler effect. Another group has shown¹⁰ that, even when loaded with cloud of relatively hot ions, such a linear trap has an advantage over the usual quadrupole trap: for a given tolerable level of second-order Doppler effect, a greater number of ions can be stored in the linear trap than in its quadrupole counterpart. □

Daniel Segal and Richard Thompson are in the Blackett Laboratories, Imperial College, Prince Consort Road, London SW7 2BZ, UK.

1. Birkel, G., Kassner, S. & Walther, H. *Nature* **357**, 310–313 (1992).
2. Itano, W. M. & Wineland, D. J. *Phys. Today* **40**(6), 34–42 (1987).
3. Diedrich, F., Bergquist, J. C., Itano, W. M. & Wineland, D. J. *Phys. Rev. Lett.* **62**, 403–406 (1989).
4. Blümel, R. *et al.* *Nature* **334**, 309–313 (1988).
5. Diedrich, F., Peik, E., Chen, J. M., Quint, W. & Walther, H. *Phys. Rev. Lett.* **59**, 2931–2934 (1987).
6. Waki, I., Kassner, S., Birkel, G. & Walther, H. *Phys. Rev. Lett.* **68**, 2007–2010 (1992).
7. Wineland, D. J., Dalibard, J. & Cohen-Tannoudji, C. *J. opt. Soc. Am.* **B9**, 32–42 (1992).
8. Hasse, R. W. & Schiffer, J. P. *Ann. Phys.* **203**, 419–448 (1990).
9. Raizen, M. G. *et al.* *J. mod. Opt.* **39**, 233–242 (1992).
10. Prestage, J. D., Tjoelker, R. L., Dick, G. J. & Maleki, L. *J. mod. Opt.* **39**, 221–232 (1992).

HMG has DNA wrapped up

David M. J. Lilley

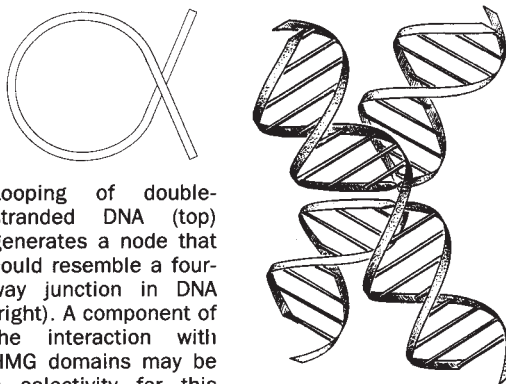
A NEW kind of eukaryotic protein domain involved in interactions with DNA is fast making its way into the literature. Known as the HMG domain, it has been discovered in a series of transcription factors, sex-determining proteins and other proteins associated with DNA in one way or another. Observations described in a flurry of papers on HMG and its analogues suggest that the domain has a major structural role in the bending or looping of DNA; such contortions are increasingly found to be required for achieving the correct conformation for transcription and various classes of DNA rearrangement. The latest of these studies, by Giese *et al.*¹, shows that the transcription factor LEF-1 (lymphoid enhancer factor), which contains a single HMG domain, can bend DNA almost back on itself.

HMG1 was discovered in the early 1970s as one of a series of acid-soluble non-histone nuclear proteins, the so-called high-mobility group proteins² (whence HMG). They turned out to be quite common, yet their function was never satisfactorily identified. HMG1 has a curious primary structure³, comprising three distinct regions. The carboxy-terminal third is extremely acidic, being composed almost exclusively of aspartate and glutamate. The remaining two-thirds is made up of two homologous repeats of an 80-amino-acid sequence — the HMG box or domain. It is this motif that is now being discovered (in one or more copies) in many DNA-binding proteins, including transcription factors such as hUBF⁴ and LEF-1⁵, the testis-determining factor SRY⁶ and yeast proteins that are involved in mating-type control^{7,8}. This leaves the questions of what the actual function of the domain might be, and how this is translated into structural terms.

From comparison of the binding sites for similar HMG-domain proteins, it seems that the structure of the DNA may be as or more important than sequence requirements. Bianchi *et al.*⁹ isolated a protein from rat liver nuclei that bound to DNA four-way junctions (see figure) constructed from synthetic oligonucleotides, and which they identified as HMG1. Isolated HMG boxes retained the structural specificity of binding¹⁰, which was found to be dependent on the DNA structure (that is, the junction) and independent of base sequence.

This intriguing observation leaves

something of a paradox; why should a protein as abundant as HMG1 in the eukaryotic nucleus exhibit structural selectivity for DNA junctions? Perhaps the junction resembles the natural substrate for these proteins that is formed in the course of wrapping or looping DNA. The four-way DNA junction is folded to generate an X-shape by the pairwise



Looping of double-stranded DNA (top) generates a node that could resemble a four-way junction in DNA (right). A component of the interaction with HMG domains may be a selectivity for this crossover structure, which has two faces where angles of approximately 120° are exposed (these are not equivalent in the four-way junction), and two of 60°. Proteins such as certain resolving enzymes appear to measure these angles¹⁸. The exact location of the protein, and the angle of DNA bending, may vary between different proteins that have recruited the HMG domain.

coaxial stacking of helical arms (see figure). The X-shape is antiparallel and right-handed, because this leads to a favourable accommodation of strands in grooves¹¹. A similar arrangement has been found in crystal structures of DNA duplexes (ref. 12, and R. E. Dickerson, personal communication) and it is clearly a natural way to pack together two DNA molecules. When a loop is formed in DNA, similar interactions might occur in the node, or crossover region, as shown in the figure, and the function of the HMG domain might be to bind such regions. Perhaps the job of HMG1 itself in chromatin is the sealing of loops in a manner reminiscent of that proposed for histone H1.

This possibility is fast becoming a probability. There have been a number of demonstrations that HMG-domain proteins can bend or supercoil DNA, and there are now two examples of HMG-domain proteins participating in site-specific recombination reactions, where they appear to bend DNA. Giese and coworkers have shown that LEF-1 bends DNA by a large angle¹ — from gel electrophoresis, they estimate the bending as 130°, which would be greater than that achieved by the catabolite activator protein CAP¹³. Significantly,

they have carried out a bend-swap experiment in which LEF-1 and its cognate binding site can substitute for bending by integration host factor at the *attP* locus in the phage λ integrase reaction. In a similar example, R. Johnson (personal communication) has shown that HMG1 itself may replace an *Escherichia coli* DNA-bending protein in the Hin-mediated recombination reaction. A further link with recombination processes comes from the isolation of another HMG protein with a putative role in immunoglobulin gene maturation, by means of screening an expression library with a DNA probe spanning the *V(D)-J* joining region¹⁴.

In each case there is a functional connection between the HMG domain and the manipulation of DNA structure. But there is also a link with DNA repair processes. In general, damage to DNA is recognized by proteins that can detect a variety of alterations, and one is left wondering how a protein such as UvrA can detect a variety of different lesions. Many agents such as psoralens, cisplatin and perhaps ultraviolet light cause changes that result in axial distortion, and this might serve as a significant component in the recognition process. In March, Lippard and coworkers¹⁵ presented the complementary DNA sequence of a protein that binds DNA containing a cisplatin adduct. At

this stage nobody should be too surprised to learn that this protein contains an HMG domain. The chemical adduct kinks the DNA by 30–40° (ref. 16), so it

1. Giese, K., Cox, J. & Grosschedl, R. *Cell* **69**, 185–195 (1992).
2. Goodwin, G. H., Sanders, C. & Johns, E. W. *Eur. J. Biochem.* **38**, 14–19 (1973).
3. Walker, J. M., Gooderham, K., Hastings, J. R. B., Mayes, E. & Johns, E. W. *FEBS Lett.* **122**, 264–270 (1980).
4. Jantzen, H.-M., Admon, A., Bell, S. P. & Tijan, R. *Nature* **344**, 830–836 (1990).
5. Travis, A., Amsterdam, A., Berlinger, C. & Grosschedl, R. *Genes Dev.* **5**, 880–894 (1991).
6. Sinclair, A. *et al.* *Nature* **346**, 240–244 (1990).
7. Kelly, M., Burke, J., Smith, M., Klar, A. & Beach, D. *EMBO J.* **7**, 1537–1547 (1988).
8. Staben, C. & Yanofsky, C. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4917–4921 (1991).
9. Bianchi, M. E., Beltrame, M. & Paonessa, G. *Science* **243**, 1056–1059 (1989).
10. Bianchi, M. E., Falciola, L., Ferrari, S. & Lilley, D. M. J. *EMBO J.* **11**, 1055–1063 (1992).
11. Murchie, A. I. H. *et al.* *Nature* **341**, 763–766 (1989).
12. Timsit, Y., Westhof, E., Fuchs, R. P. P. & Moras, D. *Nature* **341**, 459–462 (1989).
13. Schultz, S. C., Shields, G. C. & Steitz, T. A. *Science* **253**, 1001–1007 (1991).
14. Shirakata, M. *et al.* *Molec. cell. Biol.* **11**, 4528–4536 (1991).
15. Bruhn, S. L., Pil, P. M., Essigmann, J. M., Housman, D. E. & Lippard, S. J. *Proc. natn. Acad. Sci. U. S. A.* **89**, 2307–2311 (1992).
16. Rice, J. A., Crothers, D. M., Pinto, A. L. & Lippard, S. J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4158–4161 (1988).
17. Pil, P. M. & Lippard, S. J. *Science* **256**, 234–237 (1992).
18. Bhattacharya, A. *et al.* *J. molec. Biol.* **221**, 1191–1207 (1991).

seems that this may well be the means by which it is recognized by the binding protein. To underline the importance of the domain, the same group has also shown¹⁷ that HMG1 itself will selectively bind platinated DNA.

In Britain HMG is most widely known as Her Majesty's Government, whose job is to direct the affairs of state. It seems that HMG the protein domain has

an analogous influence in directing the manipulation of DNA structure. We can look forward to rapid developments in the quest to find out how widespread that influence is, and the mechanism by which it is exercised. □

David M. J. Lilley is in the Department of Biochemistry, University of Dundee, Dundee DD1 4HN, UK.

CARBON DIOXIDE DISPOSAL

The deep-six fix again

James C. Orr

As the concentration of carbon dioxide continues to rise in the atmosphere, so the temptation to find a technical fix increases. On page 318 of this issue¹, Haugan and Drange look at the feasibility of pumping the carbon dioxide in flue gases from power stations into the shallow ocean, and find that it should dissolve efficiently and be carried to ocean depths that are well isolated from the atmosphere.

Already the sea holds perhaps 38,000 billion tonnes of dissolved carbon, 65 times the amount in the atmosphere, and it is generally accepted that the ocean acts as one of the principal ultimate sinks for the additional CO₂ being produced by human activities. But can we coerce the oceans to absorb even more? One recent suggestion² was that the addition of iron — possibly a limiting micro-nutrient — to surface waters might stimulate giant algal blooms that would take up more CO₂ from the atmosphere. Unfortunately, models of even perfectly successful iron fertilization seem less than encouraging³⁻⁵.

Another possibility⁶ is to take advantage of ocean physics and chemistry by routing CO₂ directly from power plants into the ocean below its surface mixed layer. Then the ocean's density structure and circulation pattern could prevent injected CO₂ from reaching the atmosphere for about a millennium. The drawback has been that costs may be prohibitive, because CO₂ must be first scrubbed from the exhaust gases, compressed, transported to a select site, and finally pumped to depths of at least a kilometre. Haugan and Drange's contribution is to show that it may not be necessary to pump CO₂ to great depth after all. According to their model, 99 per cent of the CO₂ gas injected into the ocean dissolves before bubbles rise 150 m, while surrounding seawater sinks after becoming more dense owing to the added CO₂.

For 15 years, proposals to inject CO₂ into the ocean have taken different forms, but their goal remains the same:

to short-circuit the natural processes of dilution in the atmosphere, gas exchange at the sea surface and, most importantly, the slow mixing of the surface with the deep ocean. Deep injection sends combusted CO₂ to depth, as a gas, a liquid or mixed with water. As originally proposed⁶, CO₂-enriched water would



Could fossil-fuel-fired power stations pump CO₂ into the sea instead of the atmosphere? Richborough power station, Kent, UK. (Photo: Mike McQueen; Impact Photos.)

be injected at depth where large differences in the natural density field and a favourable circulation pattern could isolate it from the atmosphere for a long time. Such natural conditions in the ocean occur in only a few places, far from most power stations. Perhaps, the most appropriate site lies where dense, warm, saline water flows at depth from the Mediterranean Sea, across the bottom of the Strait of Gibraltar, and into the Atlantic Ocean, subsequently sinking to about 1,500 m and spreading throughout mid-depth waters.

Shallow injection works differently. Haugan and Drange show thermodynamically that seawater becomes more

dense after addition of large amounts of CO₂. For shallow injection, they calculate that the change in density is relatively large, up to seven times the range in the natural density field for the proposed site of deep injection at the Mediterranean outflow. They also model the transfer of CO₂ from rising bubbles into the surrounding seawater. For shallow gaseous injection at 200–300 m, most of the CO₂ dissolves before bubbles rise even 100 m.

Natural conditions necessary for shallow injection are more easily met than those for deep injection. Still, suitable sites for shallow injection must have weak ambient stratification and a mixing regime favouring renewal of seawater in the bubble zone that is rapid enough to dissolve all added CO₂ but not so rapid as to preclude sinking. Shallow injection should occur near the coast and above a sloping bottom along which the sinking bottom gravity current could then descend into the deep-ocean. If this CO₂-rich water moving along the bottom were to encounter CaCO₃-rich sediments, the ensuing reaction, forming bicarbonate, would further enhance oceanic uptake of CO₂. Therefore, it seems that shallow injection has several advantages, not least that it is easier and less costly than pumping CO₂ to greater depth further offshore.

Of course, it is not so straightforward. First, CO₂ must be separated from the exhaust of power stations and compressed. This process, before the CO₂ can even be piped seaward, would derate a power station by about 35 per cent and double its cost of construction (G. Booras, Electric Power Research Institute). Subsequent transport overland may be feasible only for near-coastal power stations because pipeline costs about \$1 million for each mile. Letting neighbouring power stations share pipelines would ease that cost.

And the shallow injection approach may have unacceptable environmental effects: the impact on local marine life is likely to be dramatic before CO₂-rich waters are dispersed. According to Haugan and Drange, local concentrations of added CO₂ could reach 1 mole per litre (400–500 times natural levels), resulting in a pH below 4 and a temperature increase of 6 °C. In such acid seawater, most organisms would die. Benthic communities in the path of the sinking current would suffer most.

Another complication may arise simply because shallow injection delivers CO₂ in gaseous form. To some extent, the sparging action brought about by

injection of gaseous CO₂ would strip other gases from the seawater. As bubbles rise, CO₂ would move rapidly from the gas phase into surrounding seawater, largely because it reacts so readily with water; oxygen and most other dissolved gases in seawater are less reactive, so they would move more easily into the bubbles and tend to stay there. Locally, substantial depletion of oxygen might ensue, perhaps resulting in anoxia, depending partially upon the rate of renewal of seawater within the bubble zone.

A further concern is the formation of clathrates. These are relatively stable substances which, in this case, are rich in both CO₂ and water and form as a solid under the elevated pressure found at depth in the ocean. Clathrates should be less likely to form with shallow injection. Some might consider clathrates an unwanted by-product of CO₂ injection because of their potential accumulation on the bottom; also, they might eventually interfere with outflow at the point of injection. On the other hand, CO₂-rich clathrates have been found to occur naturally and are relatively inert (J. O. Berning, Electric Power Research Institute). When CO₂ is injected as a liquid under high pressure, its only reaction with water appears to be formation of clathrates; therefore, local acidification, heating and oxygen depletion would all be avoided.

Those recalling the Lake Nyos disaster of 1986, in which a cloud of toxic gas killed thousands in Cameroon, may wonder whether injection of CO₂ into the ocean may make something similar possible. At Lake Nyos, CO₂ from subterranean (volcanic) sources had saturated the water at the bottom of the lake; when a disturbance caused the lake to turn over, depressurization released the dissolved gases, with tragic consequences. Perhaps a similar sudden release of CO₂ might occur from the ocean if a shallow-injection site were situated above a bottom with topographic depression in which added CO₂ could accumulate. But no build up of CO₂, or consequent release, could occur so long as the dense CO₂-rich water can escape down a sloping bottom to the deep sea, as proposed by Haugan and Drange.

Although local side effects may prove dramatic, the effect of the injected CO₂ away from the injection site will be negligible. The whole ocean will not become significantly more acidic because

of the amount of carbon (the additional few billion tonnes a year from power stations would be, literally, a drop in the ocean) and because its inherent carbonate chemistry makes it a well-buffered system. Seawater density would increase near the point of injection, owing to addition of CO₂, thereby facilitating local sinking under some circumstances; however, the increase would be negligible after some dispersion, that is, with respect to any large-scale processes in the ocean that are density related, such as ocean circulation and deep water formation.

Shallow bubbling as proposed by Haugan and Drange may prove the most practicable approach to oceanic CO₂ injection; however, before it can be

implemented appropriate injection sites must be identified and local environmental consequences will require thorough evaluation. Oceanic CO₂ injection might then prove itself as a mechanism that can partially limit rising levels of atmospheric CO₂, but it could only hope to be one part of a larger strategy, simply because many fossil-fuel-fired power stations are far from the nearest coastline. And anyway, they are not the only emitters of CO₂. Certainly, any sensible strategy must still embrace conservation and resource management. □

James C. Orr is in the Program in Atmospheric and Ocean Sciences, PO Box CN710, Sayre Hall, Princeton University, Princeton, New Jersey 08544, USA.

MOLECULAR GENETICS

Fast forward in the MHC

Jonathan Howard

THE Americas remained unpeopled until recently, native American tribes of both subcontinents being descendants of one, or more probably a few, immigrations of mongoloid peoples from the Asian land mass between 11,000 and 40,000 years ago. The subsequent fate of this restricted gene pool during migration, expansion and diversification within North and South America should provide a unique source of information about the behaviour of components of the human gene pool under natural selection and genetic drift.

The alleles of the massively polymorphic major histocompatibility complex (MHC) are an invaluable collection of markers for such a project. Two analyses of the nature and frequency of class I MHC alleles in native American tribal units, published on pages 326 and 329 of this issue^{1,2}, do not show the phenomenal allelic diversity typical of the Old World. Limited sets of alleles at all three of the class I loci of the MHC are found in varying frequency in all Amerindian tribes, regardless of provenance, from which one can conclude that this Amerindian set is a legacy of the small number of founding alleles that arrived with mongoloid immigrations.

These data confirm previous work. However, the unexpected feature of these two studies is the information they contribute about the rates and modes of evolution of class I MHC molecules themselves. Both groups were led to their novel conclusions by the observation that 'typing' antisera used to identify alleles at the most polymorphic of the human MHC loci, *HLA-B*, showed atypical behaviour when applied to cells obtained from individuals of small,

ethnically well-defined tribes of South American Indians. The implication was that the *HLA-B* molecules being recognized by these typing reagents were not identical to the canonical molecules defined in Eurasian populations. Sequence analysis showed that all the novel *HLA-B* molecules differed from the canonical forms by small numbers of substitutions in the amino-terminal domains, known to form the groove in which antigenic peptides are bound. These regions of class I molecules accumulate polymorphic diversity^{3,4}, which (it is generally acknowledged) arises from selection pressure exercised by infectious disease through favouring peptide binding sites that can confer immune resistance to certain pathogens^{5,6}.

The significance of the Amerindian *HLA-B* variants arises from two further considerations. First, all the *HLA-B* variants found in the South American Indian tribes are genuine novelties that have not been seen in any other human population; in particular, they have not been detected in other Asian mongoloid populations. It might be argued that the variants were indeed represented in the founding immigrant stock, but are present at such low frequencies in the modern Asian mongoloids that they have not yet been recovered in what is, after all, still a very limited data set. If this were so, they would be expected to be present in all Amerindian tribes. But in a further analysis of *HLA-B* alleles from an Amerindian tribe in New Mexico, Watkins *et al.*² found only canonical sequences. The implication is that *HLA-B* has rapidly evolved sequence novelty, certainly since the Asian immigration, and probably since the entry of the

1. Haugan, P. M. & Drange, H. *Nature* **357**, 318–320 (1992).
2. Martin, J. H., Gordon, R. M. & Fitzwater, S. E. *Nature* **345**, 156–158 (1990).
3. Peng, T.-H. & Broecker, W. S. *Nature* **349**, 227–229 (1991).
4. Joos, F., Sarmiento, J. L. & Siegenthaler, U. *Nature* **349**, 772–775 (1991).
5. Sarmiento, J. L. & Orr, J. C. *Limnol. Oceanogr.* **36**, 1928–1950 (1991).
6. Marchetti, C. *Climatic Change* **1**, 59–68 (1977).

immigrant population into tropical South America.

The second consideration concerns the nature of the variation itself. The novel sequences consist of clustered substitutions, all of which can be fully accounted for by segmental exchanges between *HLA-B* alleles present in the Amerindian tribes. Although much polymorphic variation in *HLA* sequences can be explained in terms of segmental exchanges between alleles⁷, the complexity of the Old World material, from which this interpretation arose, leaves open other interpretations based on successive point mutations with selection and recombination. In addition, no segmental exchange event of the proposed type has yet been documented in a human family. The new data strongly support the idea that interallelic segmental exchange may be an important mutational mechanism in *HLA* genes. Indeed, they suggest that such exchanges are an overwhelmingly more important source of adaptive sequence novelty in class I MHC molecules than is point mutation.

Presumably, the exceptional rapidity of class I sequence evolution implied by the present data is due to the strong selection that diseases can impose on natural populations^{6,8} — arising in this case, no doubt, from contact with previously unencountered diseases and disease vectors as the immigrants moved south from the temperate into the tropical zone. While selection disposes, however, only mutation can provide the raw material. So, what is the significance of the exclusive use of segmental-exchange mutations rather than point substitutions in new, adaptively successful variants?

There appear to be two relevant variables, namely, the frequency of the two classes of mutation, and the proportion of advantageous events within each class. It is not necessary for segmental exchange to occur at an exceptionally high frequency. Indeed the results are perfectly consistent with segmental exchange occurring at an even lower frequency than point mutation, if the probability of an adaptively useful variant from the segmental exchange is much higher. Random point mutation has an obvious disadvantage over the permutation by segmental exchange of already

structurally valid coding sequence, in a molecular structure apparently designed as a linear array of polymorphic (and hence exchangeable) structural motifs matching the amino-acid side-chains of bound peptides^{9,10}.

As Haldane¹¹ pointed out, however, disease-resistance systems should acquire not only polymorphism, but also high mutation rates to generate allelic novelty at a high rate. It is an obvious specula-

tion, further encouraged by these new results, that the segmental exchanges so typical of allelic variation in class I MHC proteins represent the operation of an adaptive mechanism specialized to assist their rapid evolution. □

Jonathan Howard, AFRC Institute of Animal Physiology and Genetics Research, Babraham, Cambridge CB2 4AT, UK, is on sabbatical at EMBL, Heidelberg.

HIGH-ENERGY PHYSICS

Casting more light on light

David J. Miller

ACCORDING to classical theory, electromagnetic waves, or photons, can interact only with charged particles and so, being chargeless, are transparent to one another. But in the quantum theory of electrodynamics (QED), a photon can behave as if occasionally composed of a pair of virtual (short-lived) charged particles: a second photon can then interact with these (Fig. 1). The best-known example is the production of electron-positron pairs when γ -rays interact with the virtual photons in the strong electromagnetic fields close to atomic nuclei. Physicists gathered at San Diego recently* to compare notes on the remarkable science revealed in photon-photon collisions.

The experiments are a bonus of high-energy electron-positron colliders, and use the intense electromagnetic fields around the pointlike charges of the accelerated electrons and positrons to provide fluxes of virtual photons which couple to pairs of charged particles. Since the last workshop in 1988, there has been a lull in photon-photon ($\gamma\gamma$) physics, owing to the closure of the two most productive colliders, PETRA, at Hamburg (collision energy, about 40 gigaelectronvolts), and PEP at Stanford (34 GeV). But the lower-energy collider, CESR, at Cornell (12 GeV) has been upgraded to a much higher luminosity (the collision rate is much higher), and two new machines at higher energy — Tristan (70 GeV) at Tsukuba, Japan and the celebrated LEP (92 GeV) at CERN, Geneva — are now producing data.

CESR has already achieved a far bigger integrated luminosity than any previous machine ($>1 \text{ fb}^{-1}$; D. Coffman, Cornell) and should soon be providing definitive analyses on the production and decay of heavy mesons, including those comprising a charmed quark-antiquark pair. The big question here is how strongly the different mesons couple to

photon pairs. If a meson is made of a quark-antiquark pair their charge couples directly to the photons (Fig. 2a), but there is great interest in systems composed of two quark-antiquark pairs (quark 'molecules') or of only the glue particles of quantum chromodynamics

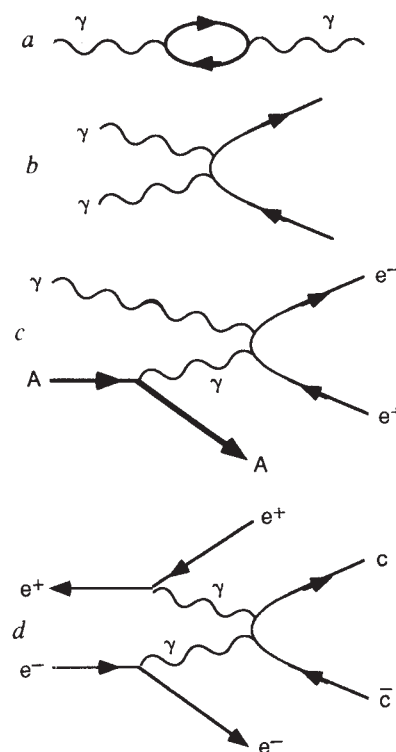


FIG.1 a, A photon (wavy line) can turn into a virtual charged particle pair (loop) in which the Heisenberg uncertainty principle allows the virtual masses to be different from the particles' real masses for a very short time. (Time progresses from left to right.) b, Two photons colliding to produce a pair of real charged particles. c, Electron-positron pair production by a real photon interacting with a virtual photon from the field of an atomic nucleus, A. d, Photon-photon collision at an electron-positron collider producing a pair of charged particles c and $\bar{c}$ which could be muon and antimuon, a quark and antiquark, a positive and negative pion, and so on.

1. Belich, M. P. *et al. Nature* **357**, 326–329 (1992).
2. Watkins, D. I. *et al. Nature* **357**, 329–333 (1992).
3. Bjorkman, P. J. *et al. Nature* **329**, 512–518 (1987).
4. Parham, P. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 4005–4009 (1987).
5. Hughes A. L. & Nei, M. *Nature* **335**, 167–170 (1988).
6. Hill, A. V. S. *et al. Nature* **352**, 595–600 (1991).
7. Parham, P. *et al. Cold Spring Harb. Symp. quant. Biol.* **54**, 529–543 (1989).
8. Howard, J. C. *Nature* **352**, 565–567 (1991).
9. Saper, M. A., Bjorkman, P. J. & Wiley, D. C. *J. molec. Biol.* **219**, 277–319 (1991).
10. Falk, K. *et al. Nature* **353**, 321–325 (1991).
11. Haldane, J. B. S. *La Ricerca Scientifica* **19**, Suppl. 325–334 (1949).

*9th International Workshop on Photon-Photon Collisions, San Diego, 22–26 March 1992.

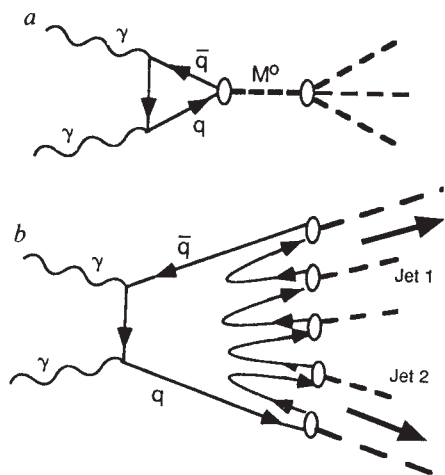


FIG. 2 *a*, Two-photon interaction via a virtual quark-antiquark loop to produce a neutral meson resonance M^0 , which in turn decays, for example, into three lighter mesons — the dotted outgoing lines. *b*, Fragmentation of the fast outgoing quark-antiquark pair into jets of mesons (see D. J. Miller *Nature* **349**, 379–387; 1991). Final-state mesons may include π , η , η' , ρ , K , $\bar{K}$ and so on.

(glueballs). Gluons are neutral, so glueballs should have negligible couplings to photon pairs. New results from the low-energy machines confirm that the two most likely candidate glueball states — the $\eta(1440)$ meson, formerly called ι and the $f_2(1720)$, formerly θ — have very weak couplings to $\gamma\gamma$, but are copiously produced in gluon-rich processes such as ψ/J decay. If the glueballs are proved to exist it will be another success for the quantum chromodynamic theory (QCD), one of the cornerstones of the standard model of particle physics.

Progress is being made in understanding how photon-photon collisions can produce pairs of mesons. This is good old-fashioned 'strong-interaction theory', using methods developed long before QCD appeared. New results on $\pi^0\pi^0$ pairs from the Crystal Ball detector at DORIS (H. Bienlein, CERN) fit well to the known $\pi\pi$ scattering amplitudes (reported by D. Morgan, Rutherford Appleton Laboratory), where the initial coupling to two photons comes from a $\pi^+\pi^-$ pair in the normal diagram (Fig. 1d). It has even been suggested that rescattering of $\pi^+\pi^-$ can account for the very high rate of production of pairs of 'vector' mesons such as $\rho^+\rho^-$ seen in many experiments (confirmed by new data from ARGUS at DORIS).

Experiments at TRISTAN (AMY and TOPAZ) and LEP (OPAL, ALEPH and DELPHI) are using their higher energies to test QCD theory at short distances in processes like that of Fig. 2*b* in which very energetic jets of particles are produced in the most violent of photon-photon collisions. Some collaborations

also reported an unexpectedly high rate for producing moderately energetic jets of mesons with their momentum transverse to the beam direction bigger than 2 GeV/c. M. Drees (DESY) presented his calculation involving gluon radiation which gives high rates for such jets; and J. Storrow (Manchester) demonstrated that a more detailed 'eikonal' calculation gives a rate only about one-third of that predicted by Drees. Present data are not precise enough to check Drees's calculations rigorously, but he predicts that these events will become increasingly prominent at future higher energy linear electron-positron colliders where they could be a serious background to studies of Higgs and W boson production and to top-quark physics.

So far all $\gamma\gamma$ physics has been done using the virtual photons around high-energy particle beams, but V. Telnov (Novosibirsk) has calculated that intense high-energy beams of real photons could be made by scattering laser light from the converging electron beams, a few millimetres from the intersection point of a linear collider. He calls it the Compton collider (Compton scattering being photon-electron scattering), and D. Borden (University of California, Santa Barbara) presented a survey of the unique physics programme which would be possible with it. The most exciting experiment would be the production of the much-sought Higgs boson by a process similar to that in Fig. 2*a*, with the Higgs taking the place of the heavy meson and with a rate calculated from a sum over all possible kinds of virtual charged particles in the loop labelled q and $\bar{q}$ which connects it to the two photons. The Higgs, which is required to explain why most particles have a mass, is predicted to couple to all such particles with a strength proportional to that particle's real mass. This coupling will exactly compensate for the suppression factor which comes in as the real masses of the charged particles in the loop become much greater than their virtual masses. So the production rate for Higgs bosons will have a finite contribution from every kind of charged particle which can couple to it, even if the particle's real mass is much greater than we can ever produce in a collider.

Of course, the collider would need to have enough energy to produce the Higgs — and we do not yet know how much that may be because we do not know its mass — but it would then put rigorous limits on the entire spectrum of elementary charged particles up to arbitrarily high masses. □

David J. Miller is in the Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK.

DAEDALUS

Erotic quality

Even pop groups these days are urging us all to fight the spread of AIDS by using condoms. Everybody loyally keeps quiet about the fact that condoms are totally charmless objects which nobody would freely choose to use. As well as introducing a depressing air of calculation, social responsibility and clinical procedure into love-making, they make it less pleasurable. The commercial evidence is convincing. Some prostitutes can charge three times as much for intercourse without a condom. Female sexual delight can also be sabotaged. The condom revolution has probably reduced the erotic quality of life by a factor of about three.

Daedalus points out the reason. The nerves of erotic pleasure are stimulated by vibration — this is why sexual vibrators work. In normal intercourse, the vibrations arise from slip-stick friction, subtly evoked by the mechanical compliance of the genital tissues and their natural viscoelastic lubrication. A condom, with its synthetic lubricant and heavily damped rubber, deadens these vibrations and quenches sexual pleasure.

So Daedalus is inventing a condom with its own built-in slip-stick friction. He remarks that no stress-strain curve is truly smooth. On a fine enough scale, every material stretches in little jerks, as individual dislocation sources are activated or molecular segments step from one site to another. DREADCO's chemists are devising a rubber which stretches in jerks large enough to be perceptible. It is loaded with tiny pairs or clumps of magnetized particles. On stretching the rubber, the clumps snap apart one by one as the local tension exceeds their force of attraction. They snap back together in reverse sequence when it is relaxed.

A coarse-grained ionomeric rubber, with clusters of mutually-attracted ion-pairs that also snap apart and snap back in this way, might also vibrate powerfully under stretching or deformation. When one of these rubbers has achieved the required amplitude and frequency characteristics, it will be formed into DREADCO's 'Passionizer' condoms. They should take the market by storm.

A Passionizer will enhance sex from the very beginning. Even putting it on will be a beguiling sensation. As it stretches and relaxes in service, vibrating subtly through every change of loading, it will lyrically enhance the pleasure of both parties. It will do wonders for those nervous or ill-acquainted lovers in whom the need to use a condom is itself powerfully off-putting. The seal of commercial success will come when prostitutes start charging more for intercourse using a Passionizer.

David Jones

Geminga: new period, old γ -rays

SIR — It was there all the time. The ROSAT results of Halpern and Holt reported in last week's issue¹ and the EGRET results of Bertsch *et al.* on page 306 of this issue² finally clinch the identification of the Einstein X-ray source 1E0630+178 (ref. 3) with the high-energy (greater than tens of MeV) γ -ray source Geminga, discovered by the satellite SAS-II and observed in more detail during the COS-B mission. Because of the paucity of γ -ray photons, it is unrealistic to search for unknown periodicities in the sub-second period range, and thus neither SAS-II nor COS-B could find any convincing time signature from the source. Armed, however, with the ROSAT and EGRET findings (the former known to one of us as referee and the second to both of us as friends), it made sense to search through the existing COS-B database⁴.

The table gives a journal of the COS-B observations when Geminga was at a reasonable distance from the pointing directions. For each observation, γ -ray photons with energies >50 MeV, and with the standard photon quality parameters used for pulsar search⁴, were selected within a few-degree cone from the source, resulting in the numbers of photons given in the table. Their barycentric arrival times were then folded in 5,000 independent steps of 1×10^{-9} each, over a period interval from 0.23709000 to 0.23709500 s, predicted on the basis of a fixed P , taken for simplic-

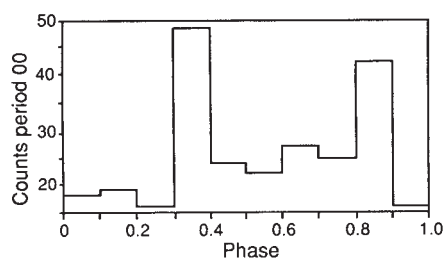


FIG. 1 Geminga γ -ray 10-bin light curve for COS-B observing period 00. The γ -rays per bin are 18, 19, 16, 49, 24, 22, 27, 25, 43, 16. Numbers for three other observing periods — 14, 39 and 64, respectively — are:

40, 37, 50, 91, 51, 38, 63, 55, 75, 64
43, 33, 31, 83, 27, 23, 36, 27, 63, 26
33, 34, 27, 70, 44, 47, 41, 40, 75, 36.

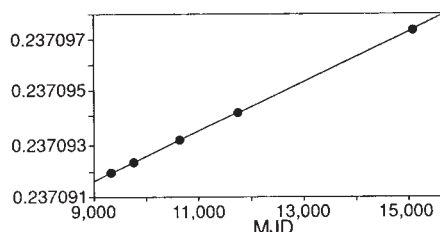


FIG. 2 Period history of Geminga, showing the four (1975–82) COS-B points and the 1991 ROSAT one. Note that each data point is enlarged for visibility, the associated error being significantly smaller. The fit regression coefficient is 0.999998 , and the derived $\dot{P}$ is $1.099 \pm 0.001 \times 10^{-14} \text{ s s}^{-1}$. Fit residuals are within ± 2.5 ns, close to the folding independent step. MJD, modified Julian day.

ity, as $1 \times 10^{-14} \text{ s s}^{-1}$. Consistently significant light curves, very reminiscent of the Crab and Vela pulsars, with two peaks separated by 0.5 in phase, were found for observations 0, 14, 39 and 64 (Fig. 1). The period values maximize the χ^2 in each individual observation. Observation 54 does not show a significant effect at the predicted period. There appears to be no obvious explanation for this, other than to note that 54 is the observation where COS-B had its lowest efficiency (47%) and the lowest photon number (224) from Geminga. According to Wills *et al.*⁵, this is the observation which also yielded the worst Crab pulsar light curve.

The secular evolution of the Geminga period can then be inferred from Fig. 2, showing the four COS-B points (1975–82) together with the ROSAT one. Over 16 years, the data fit exceptionally well to a straight line, yielding $\dot{P} = 1.099 \pm 0.001 \times 10^{-14} \text{ s s}^{-1}$ with fit residuals comparable to the precision of our period search.

The 1991 observations, supported by the old COS-B data, point to Geminga being a rotating neutron star with timing parameters that compare well to those of the population of radio pulsars. Standard formulae yield $B = 1.5 \times 10^{12} \text{ G}$, an age of 3.7×10^5 years and $\dot{E} = 3.2 \times 10^{34} \text{ erg s}^{-1}$. For a Vela-like γ -ray production efficiency, this places the source at ≤ 40 pc from Earth.

It is thus likely that Geminga, up to a

short while ago the brightest unidentified γ -ray source in the sky, is the nearest example of a magnetized, rotating, middle-age neutron star. Despite repeated searches, however, Geminga is not a radio pulsar, at least so far: the vast majority of its energy flux is emitted in γ -rays ($2 \times 10^{-9} \text{ erg s}^{-1} \text{ cm}^2$), with a fraction (10^{-3}) in X-rays and an extremely faint optical counterpart^{6,7}. As successfully shown for Vela^{8,9} (at 450 pc), a search for proper motion of G', the proposed optical counterpart^{6,7}, if indeed at a few tens of parsecs, could be pursued.

To gain clues about isolated neutron stars, which are not radio pulsars, it is important to find other, similar objects. In this context, we draw attention to 1E 1257.4–5209 (ref. 10), an HRI Einstein X-ray source with no optical counterpart ($L_x/L_{\text{opt}} > 1,000$), at the centre of a $> 10^4$ -year-old SNR, soon to be observed by ROSAT, but not yet observed in γ -rays.

G. F. Bignami*

P. A. Caraveo

Istituto di Fisica Cosmica del CNR,
15 via Bassini, 20133 Milano, Italy

* Dipartimento Ingegneria Industriale,
Università di Cassino, Italy

1. Halpern, J. P. & Holt, S. S. *Nature* **357**, 222–224 (1992).
2. Bertsch, D. L. *et al.* *Nature* **357**, 306–307 (1992).
3. Bignami, G. F., Caraveo, P. A. & Lamb, R. C. *Astrophys. J. Lett.* **272**, L9 (1983).
4. Mayer-Hasselwander, H. A. *Explanatory Supplement to the Final COS-B Database* (1985).
5. Wills, R. D. *et al.* (Caravane collaboration) *Nature* **296**, 723 (1982).
6. Bignami, G. F., Caraveo, P. A., Paul, J. A., Salotti, L. & Vigroux, L. *Astrophys. J.* **319**, 359 (1987).
7. Halpern, J. H. & Tytler, D. *Astrophys. J.* **330**, 201 (1988).
8. Bailes, M., Manchester, R. N., Kesteven, M. J., Norris, R. P. & Reynolds, J. E. *Astrophys. J.* **343**, L53 (1989).
9. Ogelman, H., Koch-Miramond, L. & Auriere, M. *Astrophys. J.* **342**, L83 (1989).
10. Bignami, G. F., Caraveo, P. A. & Mereghetti, S. *Astrophys. J. Lett.* **389**, L67 (1992).

Big Bang contd. . .

SIR — There are several points in the article by Peebles *et al.*¹ on the standard relativistic hot Big Bang cosmology with which we could take issue, but here we confine ourselves to the most important ones.

Contrary to what Peebles *et al.* claim, if a thermalizing agent absorbing and re-emitting microwave radiation operates after clusters of galaxies are formed, there is no difficulty in explaining the unique temperature and extreme smoothness of the background. Their claim referred to the classical steady-state model, which was not the model discussed by us².

The correct argument runs as follows: as an initially non-thermodynamic radiation field is thermalized to black-body form, irregularities in the distribution of the thermalizing agent are lost, becom-

Obs. no.	Date	Pointing direction (l, b)	Relative efficiency	Photons	Period	Red χ^2 (9 d.o.f.)
00	75/08/17	186°, – 6°	1.00	259	0.237091993	4.95
	75/09/17					
14	76/09/30	195°, + 4°	0.97	564	0.237092385	5.32
	76/11/02					
39	79/02/22	190°, 0°	0.69	392	0.237093213	9.49
	79/04/03					
54	80/09/04	188°, – 3°	0.47	224	—	—
	80/10/17					
64	82/02/18	190°, 0°	0.55	447	0.237094249	5.57
	82/04/25					

ing undetectable when the black-body spectrum is reached. Suppose the initial background before the operation of the thermalizing agent is mottled by the piecewise structure of our model, so that an observer viewing the initial situation in the absence of thermalization would find the background to be variable, like a chessboard seen with the sides of its squares subtending an angle of a few degrees. Now place an isotropically scattering screen of optical depth τ between the observer and the chessboard. Only three or four scatterings are sufficient to reduce the contrast in the main radiation field well below the observational limit on the smoothness of the microwave background. For $\tau \gg 1$ the contrast of the chessboard squares is essentially completely lost in the scattered field. All that remains of the original mottling is then carried by the small fraction, $\exp - \tau$, of the radiation which penetrates the screen without being scattered. Even an initially high contrast is effectively lost when $\tau \approx 10$ or more. Such an opacity, when produced by absorption and re-emission rather than by simple scattering, also maintains the black-body spectrum emerging from each of the inflationary pieces.

Writing $\tau_0 H_0/c$ for the present-day optical depth per unit path length, the optical depth τ extending to the look-back redshift z when the overlap of pieces is considered to have taken place is given to sufficient accuracy by $\tau \approx \tau_0 (1+z)^{3/2}$. With $\tau_0 \approx 1$ (which is not contradicted by any data) we have $\tau \approx 10$ for a lookback redshift $z \approx 4$. The overlap of pieces at this redshift can be considered as the transition between cosmology and astrophysics. From an observational point of view, it is interesting that the transition can be at a lookback redshift which is accessible to the telescopes of today and tomorrow. This contrasts favourably to the situation in the standard model where the conditions responsible for astrophysics are claimed to lie in an early universe far outside the observable range of tomorrow as well as today.

There is nothing in our model that should be objectionable to the particle physicists, since the detailed physical conditions applying to each inflationary piece are no different from what is proposed in the standard model. The situation is indeed improved on in the same way that it is for astrophysics, by bringing the connection between theory and observation within a practical range. We do not believe that it is possible to advance science profitably when the gap between theoretical speculation and observation/experiment becomes too wide, as we feel it has done in cosmology over the past two decades, and as it is tending to do now in particle physics, for

what then happens is that scientists are impelled to chase after chimaeras instead of real beasts. Effects are expected but not found, which has indeed become the rule rather than the exception. Of the many failures are the failure to explain galaxy formation, the failure to find fingerprints of galaxy formation in the microwave background, the failure to identify 'missing mass', the age problem, the baryon-to-photon problem, the top quark mass problem, the Higgs boson mass problem, the inflationary switch-on problem, the inflationary switch-off problem and, ultimately, the origin problem. In view of this negative record, we find the enthusiastic claims made by Peebles *et al.* surprising. The time has surely come to open doors, not to seek to close them by attaching words like 'standard' and 'mature' to theories that, judged from their continuing non-performance, are inadequate.

Finally, a comment about the evidence for non-cosmological redshifts of quasi-stellar and related objects which we have taken into account in our model. Peebles *et al.* cast doubt on this argument by invoking as others have "subjective selection effects and lack of rigorous control in the surveys". We ask them and the reader only to look without prejudice, particularly at our extensive work³ addressing all these questions.

H. C. Arp

Max-Planck Institut für Astrophysik,
8046 Garching bei München,
Germany

G. Burbidge

Center for Astrophysics and Space
Sciences,
University of California, San Diego,
La Jolla, California 92093, USA

F. Hoyle

School of Mathematics,
University of Wales,
Cardiff CF2 4AG, UK

J. V. Narlikar

Inter-University Centre for Astronomy
and Astrophysics,
Ganeshkhind, Pune 411007, India

PEEBLES *ET AL.* REPLY — A central point of our review article is that the measured properties of the thermal cosmic background radiation seem to require that the Universe has expanded from a state denser and hotter than it is now¹. We are glad that Arp *et al.* have come to accept this conclusion.

Another central point of our article is

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.

that the observed abundances of the lightest elements seem to require that the expansion and cooling of the Universe traces back to an expansion factor (redshift) greater than about 10^{10} . The extensive unsuccessful efforts to find a way around this conclusion can be traced from the references in our article. Arp *et al.* do not comment on this evidence; if they can see a sensible way around it the cosmology community surely would respond with great interest.

Arp *et al.* correctly note that in the standard relativistic hot expanding cosmological model there are many open problems: we cannot explain how the galaxies formed or even the nature of their dominant component, the massive dark halos. The many ideas for how these problems might be resolved is the focus of much of contemporary research in cosmology. Perhaps this research will lead to an impasse showing there is something wrong with the framework provided by the standard cosmological model, but as we emphasized¹, that certainly has not happened yet. No well-established phenomenon or problem is known to contradict the standard model.

The standard model traces back to a formal singularity from which one presumes we are rescued by new physics. That may turn out to be inflation, but inflation is not part of the standard model because there are not yet reasonably convincing tests of the idea. Perhaps the new physics will go some other way. We mentioned¹ the idea first put forward by Hoyle and Narlikar in 1966, and noted by Arp *et al.*, that material may be created in bursts between which the Universe expands as in the standard model. If the bursts are hot and dense enough they can produce the conditions for thermalization of the cosmic background radiation and for light element production. It thus appears to us that when the parameters in the model discussed by Arp *et al.* are adjusted to fit the observations their picture reduces to what is commonly known as the hot Big Bang model.

P. J. E. Peebles

E. C. Turner

Departments of Physics and
Astrophysical Sciences,
Princeton University,
Princeton, New Jersey 08544, USA

D. N. Schramm

R. G. Kron

Department of Astronomy and
Astrophysics and Yerkes Observatory,
University of Chicago,
Illinois 60637, USA

1. Peebles, P. J. E., Schramm, D. N., Turner, E. L. & Kron R. G. *Nature* **352**, 769–776 (1991).

2. Arp, H. C., Burbidge, G., Hoyle, F., Narlikar, J. V. & Wickramasinghe, N. C. *Nature* **346**, 807–812 (1990).

3. Burbidge, G., Hewitt, A., Narlikar, J. V. & Das Gupta, P., *Astrophys. J. Suppl.* **74**, 675–730 (1990).

Much ado about species

Jerry A. Coyne

The Units of Selection: Essays on the Nature of Species. Edited by Mark Ereshefsky. MIT Press: 1992. Pp. 405. \$60, £44.95 (hbk); \$33, £24.75 (pbk).

No area of evolutionary biology has been more beset by semantic and philosophical squabbles than the study of speciation. The difficulty of understanding such a slow historical process has repeatedly driven scientists out of their laboratories and into the arms of philosophy. From this union has sprung a bloated, quasi-philosophical literature about whether species exist, what they are and whether they differ from more arbitrary categories such as genera or families. There is some point to this debate; after all, it is hard to study speciation without a reasonable notion of its products. For most evolutionists, however, these matters were settled 50 years ago by Ernst Mayr, who proposed the widely accepted "biological species concept", defining species as objectively real groups of interbreeding populations, separated from other such groups by reproductive isolating barriers such as sterility and mate discrimination. But the debate continues, fuelled by a new generation of philosophers and a curious animosity between evolutionists and systematists. Their recent writings are collected in *The Units of Evolution*, which contains 18 previously published papers aimed at both biologists and philosophers of science.

The first series of papers, by working biologists, deals largely with the reality and definition of species. Most authors (with the exception of Robert Sokal,

Theodore Crovello, Brent Mishler and Michael Donoghue) admit that species are real and not arbitrary groups demarcated by humans. There is, however, no agreement about the nature of species, save that it is not adequately described by the biological species concept. Once again we hear the standard catalogue of objections to Mayr's definition: it fails to encompass asexual taxa, geographically separated populations, plants (which are supposed to hybridize pervasively), species that are paraphyletic and different stages of a single evolving lineage. Many of these objections are based on a misunderstanding of the concept's purpose — to describe the discontinuities between groups of sexually reproducing organisms existing in one locality — and have been repeatedly addressed by Mayr and others over the years. Other objections are not scientifically well-founded. Many authors, for example, assert that asexual 'species' are just as distinct as sexual ones, so that reproductive isolation cannot be part of a species concept. Unfortunately, there is little support for this view. First, very few groups are truly asexual. Most, like bacteria, actually have some form of genetic recombination, which can be fairly extensive. Other supposedly asexual groups, such as bdelloid rotifers and Fungi Imperfecta, may actually have sex on the sly, with the reproductive phases not yet identified. But the real problem is the lack of

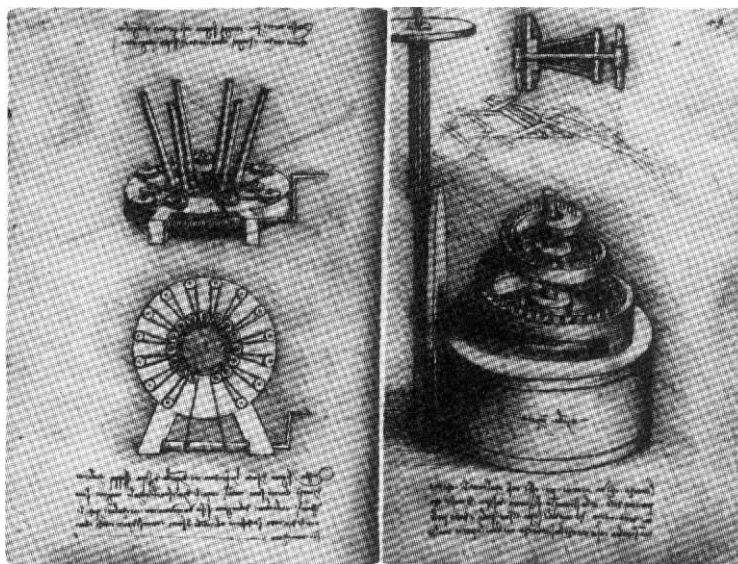
any work demonstrating that asexual and sexual species are equally distinct. Given their different evolutionary dynamics, these two groups may never adhere to a single species concept.

Similarly, the messiness of some plant 'species' may not reflect a failure of the biological species concept, but only the ability of plant populations to undergo substantial morphological evolution without becoming reproductively isolated. Many 'hybridizing' plant species may represent only a single polytypic species differentiated in a few characters by strong habitat selection, a possibility that has been largely ignored.

To replace the biological species concept, the authors proffer nearly a dozen new species concepts, some of them quite ingenious. So, for example, species are defined as phenotypically similar populations (Sokal and Crovello), lineages occupying ecological niches different from those of other lineages (Leigh Van Valen), phenotypically diagnosable groups of organisms having a parental pattern of ancestry and descent (Joel Cracraft), populations sharing a common fertilization system (Hugh Paterson) and populations of demographically equivalent organisms sharing phenotypic cohesion mechanisms (Alan Templeton). Other authors, such as Philip Kitcher, Mishler and Donoghue, offer many-part definitions to cover all existing organisms.

For sexually reproducing organisms living in the same area, nearly all of these concepts identify the same species as does the biological species concept. But these alternative concepts suffer from either the same problems as Mayr's or their own unique problems. Ecological and cohesion concepts are impossible to use in practice, and fail to separate

Industrious Inventor — as an engineer, Leonardo Da Vinci designed, among other things, an olive press, counterweights to make a door close automatically, candelabras, a variable-intensity table lamp, some folding furniture, locks and chests, an octagonal mirror that would multiply indefinitely the image of anyone standing inside it (he may have used this for his self-portrait), a therapeutic armchair and a crane for emptying ditches. Pictured here is his drawing of a stretching device and volute gear. The drawing is taken from Serge Bramly's biography entitled *Leonardo: The Artist and the Man*, which was first published in France in 1988 to critical and popular acclaim. The English translation of the book has just been published by Michael Joseph, price £20.



demonstrably different species with similar niches, such as introduced species that outcompete endemic ones. Concepts based on morphological differences lead to extreme splitting — with each slightly different population becoming a new species — and cannot distinguish among sibling species that are morphologically identical but reproductively isolated. (Such morphology-based definitions come mainly from systematists, who are more interested in classifying species than in understanding how they arise.) The authors snipe at one another's concepts, with some of the best criticisms coming not from biologists but from philosophers such as Michael Ghiselin and Kitcher. When the dust has settled, however, only Mayr is still on his feet, with his original concept remaining the simplest and most useful view of species. This is instinctively recognized by evolutionists, who nearly always focus on the origin of reproductive isolation when studying the origin of species.

The second series of papers, by philosophers of science, concentrates on two questions: are species the units of evolution, and do they consist of groups of individuals or are they individuals themselves? Although discussion of the first question fills many pages, the answer seems clear. Species are not the only units of evolution; natural selection can, for example, produce differences among populations of a single species. Such local adaptations must usually be as ephemeral as the populations themselves.

The sexually reproducing species, however, is an evolutionary unit in a very important sense. As a reproductive

community, it delimits the spread of a generally advantageous allele, such as one increasing fecundity by two per cent. This fact must surely explain the morphological unity of many species. Resolution of the second question, whether species are entities or merely groups of entities, is repeatedly touted as the key to understanding speciation, but it is unclear exactly what important biological questions could then be answered.

Although providing useful exercise for minds fatigued by hours at the microscope, this volume is not a major contribution to the study of speciation. It is clear that the arguments will persist for years to come, but equally clear that, like barnacles on a whale, their main effect is to retard slightly the progress of the field. Understanding speciation will ultimately require less rumination and more perspiration. □

Jerry Coyne is in the Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637, USA.

The making of the atom bomb

Rudolf Peierls

British Scientists and the Manhattan Project: The Los Alamos Years. By Ferenc Morton Szasz. *Macmillan/St Martin's Press*: 1992. Pp. 167. £40.

The Los Alamos Primer: The First Lectures on How to Build an Atom Bomb. By Robert Serber. *University of California Press*: 1992. Pp. 98. \$23.

BRITISH scientists contributed to the work of the Manhattan Project (the code-name given to the secret wartime effort in the United States to make atomic weapons) in three ways. First, they persuaded the British government, and thus indirectly the US government, that an atom bomb could be built, and that one was urgently needed as a deterrent against a possible German equivalent. Second, British teams took part, after 1943, in the bomb research at Los Alamos and in the work on electromagnetic isotope separation at Berkeley and at Oak Ridge, Tennessee, as well as acting as consultants to the Kellogg team in New York, which designed the gaseous diffusion separation plant at Oak Ridge. Finally, some groups in Britain worked on special problems for the Manhattan Project. In *British Scientists and the Manhattan Project*, Szasz mentions and commends all these contributions. But, as the subtitle of the book indicates, his real interest lies in the British mission to Los Alamos.

The secret laboratory established in Los Alamos in 1943 was without doubt an exciting place. There, in a remote corner of the New Mexico mountains, was gathered a unique group of talented researchers (including at least eight actual or future Nobel laureates) faced with the challenge of building bombs. It was especially exciting for the British participants, many of whom were new to North America, all to the New Mexican desert. By collecting a tremendous amount of source material, Szasz has succeeded in painting a vivid, albeit kaleidoscopic, picture of those years. One of his aims is to assess the value of the British contribution, but as he is not himself an expert, he has to quote various people's opinions on this question. And naturally there are various opinions.

Readers familiar with the story will notice a few errors in dates and other facts. For example, the author tells us that the Los Alamos laboratory was set up in an isolated place so that scientists could talk to one another freely without having to be compartmentalized. In fact, the purpose of isolation was to afford greater security and lessen the risk of leaks. The absence of compartmentalization was a concession to Oppenheimer after he had agreed to the isolation. Some of the other inaccuracies relate to the way of life at Los Alamos; here Szasz fills in gaps in his sources by using common sense. But security and military administration do not always operate by common sense. These slips, however, are not serious enough to distort the overall picture.

Szasz concludes with a sketch of the postwar history of US-UK atomic relations, brief résumés of the later careers of members of the British mission to Los Alamos, and an appendix containing the text of two memoranda written by Otto Robert Frisch and myself to the British government in 1940 to draw attention to the possibility of nuclear weapons. Altogether, he tells the remarkable story of the Los Alamos adventure in clear if not always elegant prose.

An interesting document from those times is *The Los Alamos Primer*. This contains the notes of a series of introductory lectures that Robert Serber gave to members of the laboratory when it started and to later arrivals. It is a clear and concise exposition of what was known at the time and what problems there were to be solved, testifying to the vision of the team of physicists who under Oppenheimer had prepared the ground for work on an atomic bomb. Although some details have of course changed since the primer was written, most of it could still serve as a first introduction today.

The primer was declassified in 1965

1992 Science Book Prizes

The winners of the 1992 Science Book Prizes were announced yesterday at the Science Museum in London. Administered by the Committee on the Public Understanding of Science and the Science Museum, and sponsored by Rhône-Poulenc, the prizes are awarded for books judged to have contributed most to the public understanding of science. Jared Diamond won the £10,000 prize for a book intended for general readership with *The Rise and Fall of the Third Chimpanzee*, in which he explores the evolutionary origins of our species, its development, special features and future. In reviewing the book in *Nature* (352, 676; 1991), Robert Attenborough described it as "enjoyable, stimulating and audacious", adding that it "confirms Diamond as an impressive scholar and popularizer". Now published in paperback by Vintage, £6.99.



Los Alamos National Laboratory/Science Photo Library

Working to rule — employees arriving at the main research building at Los Alamos.

and now makes its appearance as a published book. The text of the original document, actually notes by Edward Condon of Serber's lectures, occupies less than 30 pages. It is interspersed with explanatory notes, aimed presumably at readers with only a rudimentary knowledge of physics, if any at all. At the beginning Serber provides an estimate of neutron multiplication using the diffusion equation (which assumes that the mean free path is small compared to the dimensions of the system). This is in fact a very poor approximation for the purpose. He then goes on to develop somewhat laboriously a better method. This is surprising, because at the time there was an exact calculation published in the open scientific literature.

Serber has now added an autobiographical foreword in which he tells how

he came to work with Oppenheimer and about the early research into atomic energy before the days of Los Alamos. A historical introduction is provided by Richard Rhodes (the author of *The Making of the Atomic Bomb*, Simon and Schuster, 1986). About the only similarity between the primer and Szasz's book is that the Frisch-Peierls memoranda, which were not known at Los Alamos when the lectures were given, are reproduced in an appendix. Nearly 50 years on, the primer still makes interesting reading, and will be particularly appreciated by historians and others studying the development of nuclear physics. □

Sir Rudolf Peierls, emeritus professor of theoretical physics in the University of Oxford, may be contacted at Flat B, 2 Northmoor Road, Oxford OX2 6UP, UK.

Completing the equations

Andrew Warwick

The Maxwellians. By Bruce J. Hunt. Cornell University Press: 1991. Pp. 266. \$34.95.

THE history of British electromagnetic theory after James Clerk Maxwell has, until recently, attracted little attention from historians of science. Received histories of physics tend to be rather linear, pausing in Britain in the nineteenth century just long enough to chronicle the contributions of Michael Faraday and Maxwell, before moving on to Germany in anticipation of relativity and quantum theory. In these histories, Maxwell is most remembered for his fundamental equations of electromagnetism and his prediction of the existence of electro-

magnetic waves. But, as Bruce Hunt notes, Maxwell's fullest statement of his theory (set down in his *Treatise on Electricity and Magnetism* of 1873) contains neither the four famous 'Maxwell's equations' nor even a hint at how electromagnetic waves might be produced or detected. In *The Maxwellians*, Hunt sets out to describe how, following Maxwell's premature death in 1879, what would become known as 'Maxwell's theory' was developed in late-Victorian Britain.

Hunt's account of the emergence of electromagnetic field theory is refreshingly different from those offered by other authors. He neither focuses excessively on the achievement of a single individual nor offers a general survey of contributors. Rather, he draws on a wealth of unpublished manuscripts

and correspondence to chart the interaction between three physicists, George FitzGerald, Oliver Lodge and Oliver Heaviside, who were gradually drawn together during the 1880s by a common interest in Maxwell's *Treatise*. Hunt argues that it was due largely to the efforts of these men — the 'Maxwellians', as he appropriately dubs them — that a Maxwellian orthodoxy emerged in Britain. It was, for example, FitzGerald who turned Maxwell's brief remarks on the electromagnetic theory of light into a comprehensive account of optical reflection and refraction. Similarly, it was Heaviside who recast the maze of analysis in Maxwell's *Treatise* into the four powerful vector equations associated with 'Maxwell's theory', thereby asserting the primacy of fields of force (rather than charges and currents) in Maxwellian electromagnetics.

Hunt's insistence on the collaborative nature of the Maxwellian enterprise does not prevent him from giving perceptive biographical accounts of the main characters and, crucially, investigating the origin of their interest in Maxwell's extraordinary book. Indeed, their notable differences of circumstance, ability and temperament, made obvious by comparison of their biographies, reinforces the need for an explanation of how these men's different theoretical interests eventually blended into a common orthodoxy. Hunt argues that a consensus was achieved (albeit relatively briefly) through personal contact and intense correspondence. In this context, Hunt also makes effective use of another main theme of his book — the impetus and material foundation given to the Maxwellian research programme by the burgeoning electrotechnology industry. Lodge and Heaviside, for example, found their first important point of common theoretical interest in their respective attempts to understand the workings of the submarine telegraph and the lighting conductor.

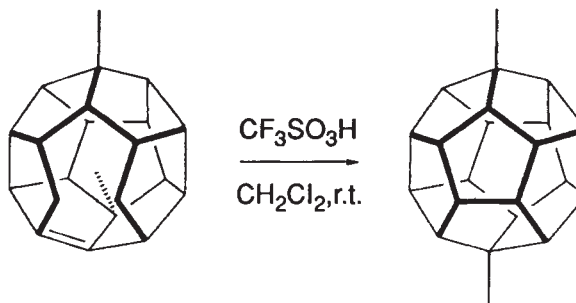
It is impossible here to do justice to Hunt's rich and meticulously documented study, but some highlights are especially worthy of note. For example, Lodge was apparently the first physicist to propose that electromagnetic radiation could be artificially generated, and to design a practical apparatus for the purpose. Hunt discusses Lodge's early correspondence with FitzGerald on this matter and describes how FitzGerald threw Lodge off the scent by showing (using Maxwell's own equations) that it was impossible to generate electromagnetic radiation artificially. FitzGerald eventually realized his error and Lodge subsequently came very close to becoming the first physicist to demonstrate the existence of electromagnetic waves in free space.

Hunt's story reaches its climax at the momentous 1888 meeting of the British Association in Bath. It was here that FitzGerald made the sensational announcement that the German physicist Heinrich Hertz had succeeded in generating and detecting electromagnetic waves. The widespread excitement produced by Hertz's experiments brought Maxwell's *Treatise* to the fore in physics and further cemented the growing friendship between its leading British interpreters. During the six years that followed — a period referred to by Hunt as the "Maxwellian heyday" — FitzGerald, Lodge and Heaviside worked in close collaboration to popularize and establish Maxwellian electromagnetic theory. One of the many pieces of Maxwellian physics from this period discussed by Hunt is the now infamous 'contraction hypothesis' proposed by FitzGerald in 1889 to account for the null result of the Michelson-Morley ether-drift experiment. The sensitive documentation of this much abused episode in the history of physics is one of the high points of the book, and should finally lay to rest discussion of the influence of this hypothesis in Einstein's subsequent work on relativity.

The Maxwellians is a remarkable achievement for several reasons. First, Hunt is completely familiar with the enormous archive (correspondence and manuscripts) of the Maxwellians; yet he manages to do justice to this literature without overly frequent, or overly long, quotation from his sources. Second, the book is exemplary of the recent trend among historians of science towards placing scientific achievements in a broader historical context. As Hunt makes clear, Maxwellian electromagnetic theory was intimately linked to the Victorian culture from which it sprang and cannot be understood in isolation. Finally, and perhaps most impressively, Hunt combines the highest level of professional historical scholarship with a narrative that is lively and compelling throughout. The only technical material over which a reader without a degree in physical sciences might stumble (Heaviside's route to 'Maxwell's equations') is conveniently relegated to an appendix. These considerable virtues make the book compulsory reading for the professional historian of modern science, and invaluable (and very welcome) as a teaching aid for students working in the field. Anyone interested in the history of physics or electrical engineering will surely find it fascinating. □

Andrew Warwick is in the Department of History and Philosophy of Science, University of Cambridge, Cambridge CB2 3RH, UK.

Chemical cookbook



Recipes for putting together just about any organic molecule from the simplest alkene to the extravagant edifice of dodecahedrane (a derivative of which is pictured here) can be found in the nine-volume series *Comprehensive Organic Synthesis: Selectivity, Strategy and Efficiency in Modern Organic Chemistry* (edited by B.M. Trost and Ian Fleming). The volumes are organized according to the bonds being formed or broken — creating carbon-carbon σ linkages, for example, or adding to C-C π bonds. As mechanistic details are provided for many of the reactions, the series is likely to be a useful reference source for research as well as an exhaustive organic 'cookbook'. Published by Pergamon Press, the complete set, including cumulative indexes, is priced at £1,950, \$3,900. **P.B.**

Thinly spread

Carl V. Thompson

The Materials Science of Thin Films. By Milton Ohring. Academic: 1991. Pp. 704. \$69.95, £46.50.

MATERIALS science is hot. Advanced materials processing, increasingly recognized as the key to the development of new or improved technologies, has spurred new interdisciplinary research-and-development groups in industry and academic institutions, and plenty of scientists are now working on topics in the traditional area of materials science and engineering.

But many of these researchers have no specific background in materials science, so there is a need for new, comprehensive texts and references in the field and its subdisciplines. This has been especially true for thin films, and of the several recent books on the subject, Milton Ohring's extensive volume is without doubt the best.

Ohring has worked and lectured on thin films for many years. He has drawn on this experience to produce a book that covers a wide range of subjects in a readily accessible way. Like authors of other books on integrated-circuit or thin-film processing, he deals with the 'unit operations' of processing, including chemical and physical deposition, thin-film reactions, diffusion and lithography. But unlike other authors, Ohring also goes into the technology of film deposition,

including diagrams and brief descriptions of the various types of vacuum pumps and evaporation techniques.

Along with the obligatory discussions of the electronic properties of thin films, Ohring includes sections on the magnetic, optical and mechanical properties of films, topics that are usually ignored in other texts. There are also chapters on film characterization and film modification techniques such as ion bombardment and laser annealing. The number of subjects covered is impressive.

But although its encyclopaedic nature is one of the book's strengths, it also contributes to its main flaws. Thin-film research is a fast-moving field, and it would be difficult for anyone to provide a thorough and up-to-date treatment of all its aspects. On any given topic, Ohring often introduces only some of the key concepts, and few areas are dealt with at a depth sufficient to make a related current research paper accessible, although references are often given to more advanced treatments and recent reviews. When there are no such references, however, the reader is left somewhat adrift.

Despite its sometimes superficial treatment, *The Materials Science of Thin Films* will be a useful introductory textbook and reference work. As a teacher and as a researcher, I shall be glad that it is on my shelf, along with the more advanced literature. □

Carl V. Thompson is in the Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Implications for climate and sea level of revised IPCC emissions scenarios

T. M. L. Wigley & S. C. B. Raper

A new set of greenhouse gas emissions scenarios has been produced by the Intergovernmental Panel on Climate Change (IPCC). Incorporating these into models that also include the effects of CO₂ fertilization, feedback from stratospheric ozone depletion and the radiative effects of sulphate aerosols yields new projections for radiative forcing of climate and for changes in global-mean temperature and sea level. Changes in temperature and sea level are predicted to be less severe than those estimated previously, but are still far beyond the limits of natural variability.

In 1990, the results of an international assessment of the greenhouse problem were published under the auspices of the Intergovernmental Panel on Climate Change, referred to below as IPCC90¹. Scenarios were given for future emissions of the major greenhouse gases, and these were interpreted in terms of future changes in greenhouse gas concentrations, global-mean temperature and global-mean sea level. Since then, there have been important scientific and methodological developments. To account for these, and as background material for the forthcoming United Nations Conference on Environment and Development, IPCC has produced an update of the 1990 assessment, IPCC92². This includes a new set of emissions scenarios^{3,4} which are more comprehensive and, methodologically, more soundly based than the earlier scenarios.

The IPCC update, however, only briefly analyses the implications of these new scenarios for future greenhouse gas concentrations, radiative forcing and global-mean temperature changes⁵; sea-level implications are not considered. Here we aim to assess the implications in detail. In so doing, we investigate not only the direct consequences of the changes in the emissions scenarios, but also how the results are affected by uncertainties in carbon cycle modelling, the negative feedback on halocarbon radiative forcing through stratospheric ozone depletion^{6,7}, and the possible negative forcing due to anthropogenic SO₂ emissions and sulphate aerosol production⁷⁻¹⁰.

The six new emissions scenarios, IPCC scenarios (IS) 92a-f, differ fundamentally from the four earlier ones (from IPCC90) in a number of ways. First, they account for new policies already implemented or proposed for controlling CO₂ emissions and halocarbon production, and allow for recent political changes. Second, and more important, they were not developed 'top down' to achieve a prescribed set of radiative forcing targets, but were based on a wide range of assumptions regarding the socio-economic factors that influence the development of emissions in the absence of new multilateral or unilateral efforts to reduce emissions^{3,4}. The new scenarios thus represent a range of possible futures all corresponding to the continuation of 'Business as Usual'. They therefore replace the single such scenario given by IPCC90. The new scenarios differ from each other because they make different assumptions regarding population growth, economic growth, technological developments, resource limitations, fuel mixes, agricultural development, the implementation or not of already-proposed policies, and so on.

The most important background assumptions for the scenarios and their emissions values for the year 2100 are given in Table 1. Radiative forcing implications and comparisons with the IPCC90 scenarios are discussed later.

The overall range of emissions projections (Table 1) may seem large given that all scenarios are based on an 'existing policies' assumption, and that IPCC90 gave only a single such scenario. Nevertheless, the range realistically reflects the extreme

uncertainties in predicting future population growth and the development of economic, agricultural and industrial activities. For CO₂ at least, as noted in ref. 4, the results are in accord with uncertainties expressed in earlier studies^{11,12}. Our main concern here, however, is not the emissions scenarios, but their implications for climate and sea level. They are taken here as the starting point for further analysis.

Concentrations and radiative forcing

To convert emissions to concentrations, gas cycle models are used. In IPCC90, only a single 'best-guess' concentration projection was given for each emissions scenario of each gas. This was obtained by averaging the results of simulations carried out independently using different models. In some cases, different models gave widely differing results¹³. These differences represent basic uncertainties in our understanding of the natural sources and sink mechanisms that determine how additional man-made emissions are reflected in concentration changes. We do not attempt to quantify these uncertainties here, although we do give two different projections for CO₂, the gas which contributes by far the most (67-96%) to future radiative forcing changes. For the other gases, we use simplified models that have been calibrated to reproduce the best-guess results from IPCC90 when used with IPCC90 lifetime data. (The simplified models reproduce the IPCC90 results well for all IPCC90 emissions scenarios¹³.) This ensures consistency with IPCC90.

To convert concentration changes to radiative forcing, we follow IPCC90 and use the relationships given in ref. 14 (Tables 2.2 and 2.4). In accord with IPCC^{7,14}, no attempt is made to include radiative forcing changes arising from possible changes in tropospheric ozone.

Carbon dioxide. The carbon cycle model we use employs a convolution integral representation of ocean uptake^{15,16}, which is based on the ocean general circulation carbon cycle model of Maier-Reimer and Hasselmann¹⁷. The ocean component is coupled to a four-box terrestrial carbon cycle¹⁸. The model has considerable flexibility in that its rate of ocean uptake can be varied in an internally consistent way by changing the decay times in the Green's function of the convolution integral^{15,16}, and the efficiency of various biological feedback processes can be altered by changing parameters in the terrestrial component of the model¹⁸.

The model is run in two modes. The first is an ocean-only mode with no feedbacks (the NFB mode (or model)). In NFB mode, the projections accurately reproduce all of the IPCC90 best guess projections¹³. In the second (FB) mode, the terrestrial biosphere is included. This introduces a variety of feedback processes into the model. The most important of these is the negative feedback due to CO₂ fertilization, the enhancement of biomass productivity caused by increasing CO₂ concentrations.

The following items apply to both cases. First, the model's mean flux of CO₂ into the ocean for the 1980s is 1.95 gigatonnes (Gt) C yr⁻¹, close to the IPCC90 best guess¹⁹ of 2 Gt C yr⁻¹. Second, as in IPCC90 (and all other projections of future CO₂ concentration changes), the relatively small sources due to oxidation of fossil-fuel-derived methane and carbon monoxide are ignored. Third, to initialize the model with 1990 conditions that are consistent with observed concentration changes before 1990, the model is first run in inverse mode from 1765 to 1990. This exposes a considerable problem that arises when using the ocean-only (NFB) form of the model.

For an inverse calculation, the model input is the observed concentration history and the output is the corresponding total emissions history, $E(t)$. When the NFB model is used in inverse mode, $E(t)$ does not agree with observationally based estimates of past emissions. For example, the implied value of $E(1990)$ is 6.2 Gt C yr⁻¹, which is noticeably less than even the lowest estimate of the sum of industrial and land-use-change emissions¹⁹. It is also much less than the 1990 value in the IS92 scenarios (7.4 Gt C yr⁻¹). This discrepancy effectively means that the NFB model fails to balance the contemporary carbon budget. The most likely explanation for this is that the model is missing a significant sink term. As a consequence, projections using the model in NFB mode will probably overestimate future concentration levels. In IPCC90, it was the NFB model that was used (with an unbalanced 1990 emissions level of 6.2 Gt C yr⁻¹). For any given emissions scenario, IPCC90 concentration projections are likely to be too high.

In FB mode, this problem is avoided by including CO₂ fertilization feedback as an additional sink term. The fertilization term is logarithmic in concentration, and its strength is determined by the CO₂ fertilization factor (β factor). As noted by IPCC92²⁰, the global-scale or ecosystem value of β is not known, although we do have some idea of its value from small-scale plant growth experiments. We can therefore leave β as a free parameter in the model. The inverse calculation can then be done with $E(1990)$ specified at any required level, and β adjusted, within realistic limits, to match this level. Achieving a contemporary balance of the carbon budget in this way means that the FB projections should be better than the NFB projections, but there are still important uncertainties.

For our FB projections, we are assuming that the 'missing sink' is explained solely by the CO₂ fertilization effect. Although there is increasing evidence that the CO₂ fertilization effect is the main factor involved in explaining the missing sink²¹, other factors could also be operating²⁰, such as enhanced plant growth due to nitrogen fertilization or increasing temperature. The strengths of other sinks need not grow at the same rate as the model-projected changes in the strength of the CO₂ fertilization sink; temperature feedback may even change sign. Balancing the contemporary carbon budget using these factors as contributors to the missing sink would lead to different CO₂ projections from those given here, although they would probably lie between our FB and NFB values.

Our FB projections are reasonably robust to other uncertainties. If both ocean flux and CO₂ fertilization are adjusted to achieve a contemporary balance (greater ocean flux requiring a smaller fertilization factor and vice versa), then, for any given emissions scenario, similar concentration projections are obtained for a wide range of combinations of ocean flux and β . In other words, the projections depend critically on whether or not the contemporary carbon budget is balanced, but they are relatively insensitive to how this balance is achieved.

CO₂ concentration projections are shown in Fig. 1 for four of the IS92 scenarios. The effect of balancing the carbon budget is to reduce future concentration changes over 1990–2100 by 11–25%. Radiative forcing changes for concentration change from C_0 to C are computed using $\Delta Q = 6.3 \ln(C/C_0)$ (ref. 14). Because of this logarithmic dependence, the percentage radiative forcing reductions due to the use of a balanced model are slightly

smaller than the concentration changes, 7–22% (see Table 1). Highest percentage reductions occur for the lowest emission case (IS92c).

Methane. Methane concentration changes are calculated using a mass balance model

$$dC/dt = E/b - C/\tau - C/\tau_s \quad (1)$$

where b is a units conversion factor, τ is the chemical lifetime and C/τ_s is a soil sink term. The present soil sink is estimated to be ~ 30 Tg CH₄ yr⁻¹ (ref. 22), corresponding to $\tau_s \approx 150$ yr. We assume that τ_s is constant, although there is some evidence to the contrary²⁰. The chemical lifetime of CH₄ is certainly not a constant, changing as the concentration of the hydroxyl radical (OH) changes. These changes, in turn, depend mainly on how the concentrations of CH₄, CO and NO_x change. In the present model, τ is assumed to be a function of CH₄ concentration and emission rates of CO, NO_x and VOCs (volatile organic compounds), the latter three quantities as given in the IS92 scenarios. The emissions of these non-greenhouse gases act as proxies for their concentrations because of their short lifetimes. The functional dependence for τ involves empirical constants that have been evaluated using the results from a number of more detailed chemical modelling experiments, including those carried out as part of the IPCC90 exercise (see ref. 23 for details).

As with CO₂, IPCC90 gave only a single 'best-guess' projection for methane concentration changes under each of the IPCC90 emissions scenarios. This was obtained by averaging the results from a number of individual models. The different models used showed fairly wide departures from the average (shown in ref. 13), indicating the uncertainty that surrounds future concentration projections. The present model has been calibrated using the IPCC90 best-guess projections, and matches these projections well²³.

In calibrating the model against IPCC90 results, the best-fit value of τ_0 was 10.1 yr, compared with the IPCC90 best estimate of 10 ± 2 yr (ref. 19). Since IPCC90, the best estimate for τ_0 has been revised upward to 11.3 yr (refs 24, 25). We therefore base

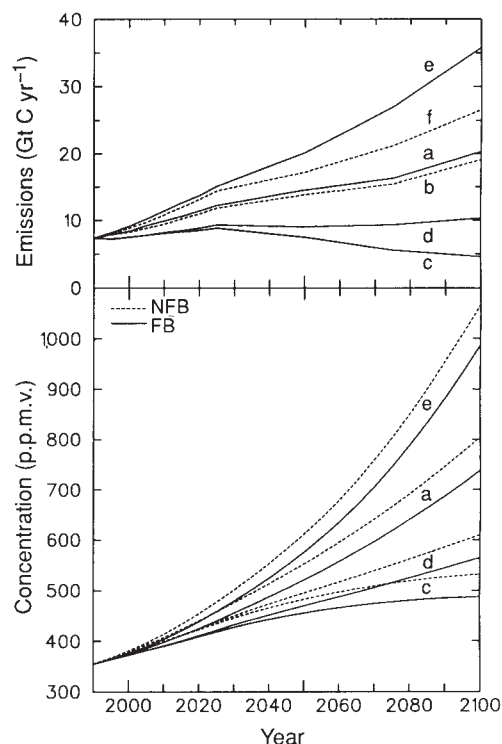


FIG. 1 IS92 emissions scenarios for CO₂ and corresponding concentration projections. Concentrations for IS92b and IS92f are not shown, to avoid confusion.

TABLE 1 Summary of basic scenario characteristics

Parameter	a	b	c	d	e	f
Socio-economic						
Population growth	Medium	Medium	Medium-low	Medium-low	Medium	Medium-High
Economic growth	Mid	Mid	Low	Low/Mid	High	Mid
Resource availability	Mid	Mid	Low	Low	High	High
Policies	Existing	Proposed	Existing	Optimistic	Existing*	Existing
Emissions (2100)†						
Fossil-fuel CO ₂ (Gt C yr ⁻¹)	20.4	19.2	4.8‡	10.4	35.9	26.6
Land-use CO ₂ (Gt C yr ⁻¹)§	-0.1	-0.1	-0.2	-0.1	-0.1	0.0
CH ₄ (Tg CH ₄ yr ⁻¹)	917	917	546	567	1,072	1,168
N ₂ O (Tg N yr ⁻¹)	17.0	16.9	13.7	14.5	19.1	19.0
SO _x (Tg S yr ⁻¹)	147	142	55	65	232	182
CFCs¶	Moderate	Low	as 'a'	Low	as 'd'	as 'a'
HCFCs*	Moderate	Moderate	as 'a'	Low	as 'd'	as 'a'
HFCs*	Moderate	Moderate	as 'a'	High	as 'd'	as 'a'
ΔQ (1990-2100) (W m⁻²)						
CO ₂ (NFB/FB)**	5.17/4.62	4.96/4.42	2.56/2.00	3.42/2.93	6.95/6.45	6.05/5.50
CH ₄ plus stratospheric H ₂ O	0.84	0.84	0.25	0.22	0.98	1.06
N ₂ O	0.27	0.27	0.17	0.19	0.32	0.31
Halocarbons (NFB/FB)††	0.40/0.36	0.33/0.34	0.40/0.36	0.23/0.40	0.23/0.40	0.40/0.36
Sulphate aerosol‡‡	-0.65	-0.61	0.20	0.10	-1.41	-0.97
Total, base case§§	6.68	6.40	3.38	4.06	8.48	7.82
Total, best guess	5.44	5.26	2.98	3.84	6.74	6.26

The socio-economic factors (upper group) represent the underlying assumptions leading to the emissions scenarios specified by IPCC92 (middle group). Only emissions for the year 2100 are shown here. The radiative forcing changes (ΔQ, lower group) are values derived here from the specified emissions.

* But with tax on fossil-fuel-derived energy (between 'd' and 'f').

† 1990 emissions: Fossil-fuel CO₂, 6.1 Gt C yr⁻¹; Land-use CO₂, 1.3 Gt C yr⁻¹; CH₄, 506 Tg CH₄ yr⁻¹; N₂O, 12.9 Tg N yr⁻¹; SO_x, 75 Tg S yr⁻¹.

‡ Peak emissions of 8.8 Gt C yr⁻¹ in 2025, then decline. Other scenarios have monotonic fossil-fuel emissions trends.

§ All scenarios are similar, but 'd' has a more rapid and 'f' a less rapid reduction in deforestation rates than the others.

|| Anthropogenic component of emissions.

¶ Qualitative information indicates the assumed rate of phase out of controlled CFCs. Moderate indicates less rapid phase out than 'low'.

* Qualitative information indicates the relative amounts of substitution by HCFCs or HFCs.

** Results for ocean-only no-feedback carbon cycle model (NFB), and balanced carbon cycle model with CO₂ fertilization feedback (FB).

†† With and without stratospheric-ozone-depletion feedback.

‡‡ Global-mean value, but forcing mainly in the Northern Hemisphere. Only the best-guess value is shown. Low and high estimates are 0.5 and 1.5 times the value shown.

§§ Base case is for NFB carbon cycle model, no ozone-depletion feedback and no sulphate aerosol forcing, thus simulating the result that would be obtained if IPCC90 methodology were used.

||| Best-guess case includes feedbacks for both CO₂ and halocarbons and uses best-guess sulphate aerosol forcing.

our projections on this value for τ_0 . This increase in the chemical lifetime (by about 10%) means that the concentration projections presented here are higher than they would have been if calculated using the models in IPCC90.

Radiative forcing changes for methane were calculated as in IPCC90¹⁴, allowing for overlap between the CH₄ and N₂O absorption spectra. We also include the contribution to radiative forcing from stratospheric water vapour produced through the oxidation of methane, calculated as in ref. 14. This contribution is uncertain, and may be less important than assumed in IPCC90 and here^{7,26}. It is, however, only a small term, amounting to a maximum of $\sim 0.3 \text{ W m}^{-2}$ by 2100. The forcing enhancement due to possible methane-derived ozone increases in the lower stratosphere²⁷ is not included. Radiative forcing information for methane is summarized in Table 1.

Nitrous oxide. Nitrous oxide concentration changes were calculated using a mass balance relationship similar to equation (1), but without a soil sink term. The N₂O case is simpler than for methane, because the lifetime can be taken as a constant (we used 132 yr, following ref. 24, slightly less than the IPCC90-recommended value of 150 yr). Radiative forcing changes were calculated using the formula given by Shine *et al.*¹⁴. Changes over 1990-2100 are given in Table 1. N₂O accounts for, at most, 0.3 W m^{-2} of the total forcing to 2100.

Halocarbons. The IS92 halocarbon emissions scenarios are extremely detailed, covering 18 individual compounds. Eight of these are possible new hydrochlorofluorocarbon (HCFC) or hydrofluorocarbon (HFC) substitutes for gases controlled under the Montreal Protocol, all of which currently have zero or near-zero emissions. In contrast, IPCC90 considered only three

halocarbons, CFC-11, CFC-12 and HCFC-22, with the latter acting as a proxy for all substitutes. Other halocarbons were considered by simply scaling up the radiative forcing of CFC-11 plus CFC-12 by 43% (ref. 14). Now, specific scenarios are given for almost all of the important gases.

Five gases that currently add to the total radiative forcing from halocarbons are, however, not covered by the IS92 scenarios CFC-13, CFC-14, CFC-116, chloroform (CHCl₃) and methylene chloride (CH₂Cl₂). We have added these to the scenarios, assuming CFC-13 emissions to be proportional to those for CFC-12, CH₂Cl₂ emissions to be proportional to HCFC-22 (because the former is used in the manufacture of the latter), CHCl₃ emissions to be proportional to methyl chloroform (CH₃CCl₃) (because these have similar uses), and emissions of CFC-14 and CFC-116 to be constant.

The latter two compounds, both of which arise mainly through the aluminium manufacturing process, are potentially important. They are not likely to be controlled under the Montreal Protocol (or its extensions) as they do not deplete stratospheric ozone, but they are strong greenhouse gases with very long lifetimes (>500 years). Furthermore, as they do not deplete ozone, they are not affected by ozone-depletion feedback (see below). With time, the radiative forcing importance of these two gases relative to most other halocarbons will increase, especially if their emissions increase.

Concentration projections for all halocarbons have been made using the same constant-lifetime mass-balance approach as for N₂O. In certain scenarios, the lifetimes of some of the gases are likely to increase noticeably between 1990 and 2100, so assuming a constant lifetime will lead to an underestimate of future

concentration levels. The net effect of lifetime variations on total halocarbon radiative forcing, however, will be small ($<0.1 \text{ W m}^{-2}$). The lifetime values used here are the latest estimates, as adopted by IPCC92⁷ and based on refs 24 (mainly) and 28.

Although some of the new lifetimes differ markedly from those given in IPCC90¹⁴, the main differences from IPCC90 arise through use of the London Amendment to the Montreal Protocol as the reference policy background. Under the IPCC90 'Business as Usual' scenario (SA90), which was based on the original protocol, the direct halocarbon radiative forcing change over 1990–2100 was $\sim 1.1 \text{ W m}^{-2}$. Under the IS92 scenarios, this change reduces to $\sim 0.4 \text{ W m}^{-2}$. The precise net forcing changes depend, however, on whether one accepts the reality of the mechanism for stratospheric ozone depletion feedback.

This feedback mechanism^{6,7} is the other potentially important new factor in assessing the overall radiative forcing contribution from halocarbons. The hypothesized effect is this. Chlorine- and bromine-containing compounds have been strongly implicated in producing stratospheric ozone loss, especially at high latitudes²⁴. The lower stratospheric cooling resulting from this loss causes a negative radiative forcing of the surface-troposphere system, which acts to offset the positive forcing due to the direct greenhouse effects of these gases. Ramaswamy *et al.*⁶ have used observed ozone losses over 1979–90 to calculate the magnitude of this effect. For the eight gases considered (CFC-11, CFC-12, CFC-113, CFC-114, CFC-115, HCFC-22, methyl chloroform and carbon tetrachloride), 80% of their total, direct, global-mean greenhouse effect is offset.

This process is a potentially important negative feedback on the radiative forcing due to halocarbons. The implications for the IS92 scenarios depend on whether or not the observed ozone losses can be attributed fully to halocarbons. If only part of the ozone loss is due to halocarbons, the negative forcing over 1979–90 would still remain, but the part attributed to, and therefore the strength of, the feedback mechanism would be reduced.

We only attempt to quantify this feedback effect in global-mean terms. The negative forcing due to ozone depletion, like ozone depletion itself, is not spatially uniform and is weighted heavily towards mid-to-high latitudes. It is clearly a gross simplification to consider only global means, but the results do give quantitative insights into the overall importance of the feedback mechanism.

First, we need to break the process down to an individual gas basis. Once again, this requires some gross simplifying assumptions: specifically, we assume that the global-mean negative forcing is linearly dependent on the global-mean stratospheric ozone loss, which in turn is assumed to be linearly dependent on the halocarbon concentration change and on the number of chlorine atoms in the molecule. (We assume that bromine atoms are five times more effective than chlorine atoms in this regard, qualitatively in accord with the large ozone depletion potentials of halons compared with CFCs²⁴.) The (negative) forcing due to ozone depletion is then

$$\Delta Q_{\text{Oz}} = \sum a N_i \Delta C_i \quad (2)$$

where a is independent of the species, and N_i and ΔC_i are the number of chlorine atoms in and the concentration change for species i . For the direct greenhouse effect (no feedback), we have

$$\Delta Q_{\text{NFB}} = \sum \alpha_i \Delta C_i \quad (3)$$

where $\alpha = \partial Q / \partial C$. Values used for α here are from ref. 14; for gases not covered by ref. 14 we have used ref. 28. For the eight species considered in ref. 6 with observed concentration changes from ref. 24, and $\Delta Q_{\text{Oz}} = 0.8 \Delta Q_{\text{NFB}}$ over 1979–90, we obtain from equations (2) and (3) $a = 0.0762 \text{ (W m}^{-2} \text{ per chlorine atom per part per } 10^{12} \text{ (p.p.t.v.))}$. We therefore evaluate past and future halocarbon radiative forcings in two ways, either using

equation (3), the standard no-feedback method, or using

$$\Delta Q_{\text{FB}} = \Delta Q_{\text{NFB}} - \Delta Q_{\text{Oz}} \quad (4)$$

to account for the feedback effect, with ΔQ_{Oz} given by equation (2).

The overall effect of the feedback is largest in the IPCC90 'Business as Usual' case (SA90), as this has much larger future CFC concentrations than any of the IS92 scenarios. For SA90, over 1990–2100, $\Delta Q_{\text{FB}} = 0.5 \text{ W m}^{-2}$ compared with $\Delta Q_{\text{NFB}} = 1.1 \text{ W m}^{-2}$. The difference is also significant for the total halocarbon forcing up to 1990, $\Delta Q_{\text{FB}} = 0.03 \text{ W m}^{-2}$ compared with $\Delta Q_{\text{NFB}} = 0.28 \text{ W m}^{-2}$. For all IS92 scenarios, however, the effect of the feedback is small, largely because the emissions of chlorine-containing compounds are much reduced from earlier (original Montreal Protocol) projections. The results over 1990–2100 are summarized in Table 1. Note that there are only three distinct halocarbon scenarios, with a, c, f and d, e being identical. For scenarios d and e, ΔQ_{NFB} is noticeably less than ΔQ_{FB} . This is because these are changes from 1990; from a zero concentration initial state, ΔQ_{NFB} must always be greater than ΔQ_{FB} . To see how the present result arises, consider ΔQ split into contributions from gases with no ozone-depletion feedback (HFCs, CFC14 and CFC116) and those with a feedback effect. The former give a positive forcing change over 1990–2100 which is the same for the NFB and FB cases. For the other gases in IS92d and e most have concentration decreases over 1990–2100. In the NFB case, they therefore contribute negatively to the forcing, whereas in the FB case (because ozone-depletion feedback counteracts the direct radiative forcing) they have a negligible overall forcing effect. The net result is that ΔQ_{NFB} is less than ΔQ_{FB} . For halocarbon radiative-forcing changes over 1990–2100, differences between scenarios become negligible when the feedback is included (see Table 1).

These feedback results have interesting implications for 'comprehensive' or multi-gas approaches to the control of greenhouse gas emissions. In the absence of feedback, the strengthening of the Montreal Protocol implied a considerable reduction in future radiative forcing, by $\sim 0.7 \text{ W m}^{-2}$ over 1990–2100 if SA90 results are compared with the IS92 scenarios after allowing for lifetime

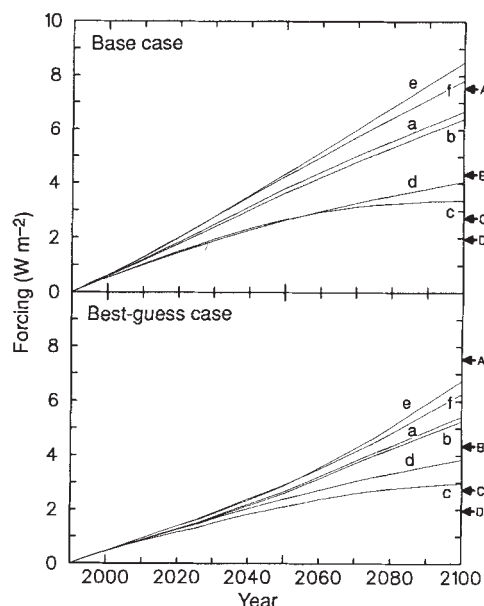


FIG. 2 Radiative forcing projections based on the IS92 emissions scenarios. The upper panel gives the base case, which follows IPCC90 in not accounting for CO_2 fertilization feedback, stratospheric-ozone-depletion feedback or sulphate aerosol forcing. The lower panel gives the best-guess case which includes both feedbacks (the effect of ozone-depletion feedback on future forcing is negligible, however) and assumes best-guess estimates of aerosol forcing. Corresponding IPCC90 scenario results for 2100 are indicated on the right (A, B, C and D).

revisions. In a comprehensive approach, this might be offset against the effects of other greenhouse gases, thereby requiring weaker restrictions on their emissions. When ozone-depletion feedback is included, the strengthening of the protocol leads to a much smaller effect on 1990–2100 radiative forcing, a reduction of only $\sim 0.1 \text{ W m}^{-2}$. This implies less latitude for easing restrictions on other greenhouse gas emissions if we are to achieve given goals of radiative forcing change.

Sulphate aerosols. The climatic importance of sulphate aerosols derived from man-made SO_2 emissions has only recently been considered seriously^{8–10,29}. There are three main effects¹⁰, a reduction in incoming short-wave radiation under clear sky conditions^{8,30}, a possible increase in cloud reflectivity due to sulphate aerosols acting as cloud condensation nuclei^{31,32} and a possible increase in cloud lifetimes³³. The first of these (the ‘direct effect’) has been best quantified by Charlson *et al.*⁸ who obtain a negative forcing of 1.07 W m^{-2} in the Northern Hemisphere for global SO_2 emissions of 71 Tg S yr^{-1} ($\sim 64 \text{ Tg S yr}^{-1}$ in the Northern Hemisphere). The stated uncertainty is a factor of two. This negative forcing is of comparable magnitude to the greenhouse-gas-related positive forcing, $\sim 2 \text{ W m}^{-2}$ over the period of significant anthropogenic SO_2 emissions, implying that aerosol changes may have noticeably offset the enhanced greenhouse effect in the Northern Hemisphere. The indirect effects through cloud albedo and lifetime changes are currently difficult to quantify reliably because of the uncertain link between changes in mass of sulphate aerosol and changes in number of cloud condensation nuclei³⁰, but they could add appreciably to the direct effect.

Because of the wide uncertainty in the aerosol forcing, we use a range of possible values. As a best-guess value for the relationship between radiative forcing and emissions under the direct effect we use the results of ref. 8 and assume that a Northern Hemisphere emission level of 64 Tg S yr^{-1} leads to a corresponding hemispheric radiative forcing of -1.07 W m^{-2} . We also assume that global emissions are always partitioned into 90% in the Northern and 10% in the Southern Hemisphere. As the estimate of the direct effect in ref. 8 assumes that emission sources are at present-day elevations, we need to account for changes in these elevations, the ‘tall stack effect’. This change has increased the fraction of SO_2 that is oxidized to sulphate, currently ~ 0.5 (ref. 8), whereas before 1950, it was probably around one-third. We account for this effect by reducing the forcing by 30% before 1950, with the reduction factor decreasing to zero in 1970. For emissions after 1990, we assume forcing to be directly proportional to the anthropogenic component of SO_2 emissions, with a 1990 global emissions level of 75 Tg S yr^{-1} (refs 4, 20; see Table 1).

For the indirect effect, we assume that it is real but small, and of current (1990) magnitude equal to 20% of the direct effect. To quantify the link between SO_2 emissions and indirect forcing more generally, we use a relationship suggested by Wigley⁹ which leads to

$$\Delta Q_i \propto \ln(1 + E/E_0) \quad (5)$$

where E_0 is the initial (natural) emissions level of sulphate precursors (taken as 42 Tg S yr^{-1} , the midpoint of the range given by IPCC92²⁰), and E is the anthropogenic emissions level. This formulation attempts to account for the nonlinear relationships between changes in initial emissions and final cloud condensation nuclei. For scenarios in which SO_2 emissions grow (a, b, e and f), it leads to a much slower rate of increase after 1990 for the indirect effect than for the direct effect. Because the indirect effect is assumed to be a minor component in 1990, and because it becomes relatively less important subsequently, the results for total radiative forcing are insensitive to uncertainties in the ΔQ_i formulation.

Finally, to account for uncertainties in the aerosol forcing we consider a range of possibilities, 50–150% of the combined aerosol forcing. As the indirect forcing is assumed to be substan-

tially less than the direct forcing, the implied uncertainty in the former is much greater than in the latter. For 1990, we deduce a global-mean aerosol forcing range of $-0.75 \pm 0.38 \text{ W m}^{-2}$. These values are much less than estimated in ref. 10, which gave best guesses of -1 W m^{-2} for both direct and indirect effects separately, and a best-guess range of -1 W m^{-2} to -2 W m^{-2} for their sum. As shown in the next section, however, and as noted by Wigley²⁹, aerosol forcing effects of this magnitude are almost certainly incompatible with observed hemispheric-mean temperature changes.

Total forcing projections. A summary of the radiative forcing projections over 1990–2100 is shown in Fig. 2. This gives two total forcing possibilities for each scenario: a base case which follows the methodology of IPCC90 and ignores CO_2 fertilization feedback, ozone-depletion feedback and aerosol effects; and a ‘best-guess’ case, in which the two feedbacks are accounted for and in which the best-guess aerosol forcing is included. Numerical values for the year 2100 are summarized in Table 1.

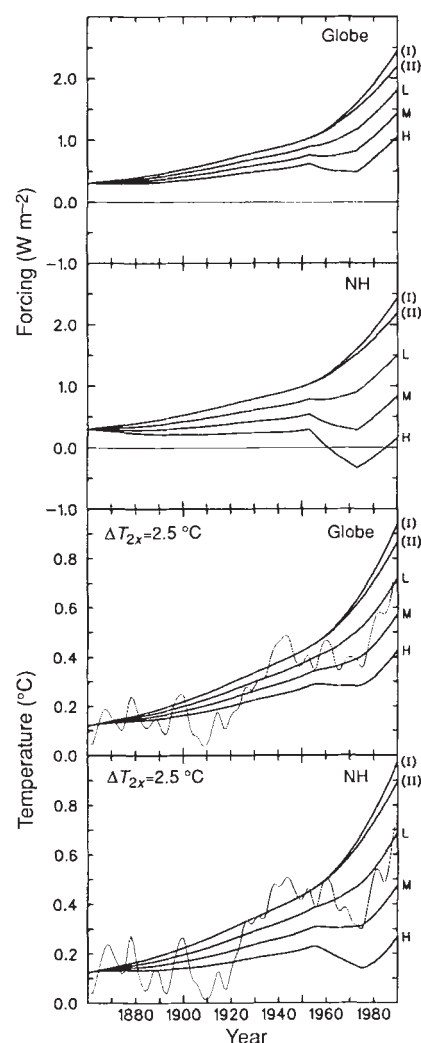


FIG. 3 Radiative forcing and temperature changes up to 1990. The panels show the effect of stratospheric-ozone-depletion feedback (compare curves I and II) and the additional negative forcing due to sulphate aerosols for low (L), middle (M) and high (H) estimates. Aerosol forcing was determined as a function of SO_2 emissions as described in the text. The anthropogenic global emissions history assumed was based on ref. 46 to 1975 and tied to the IPCC92 best-guess estimate²⁰ of 75 Tg S yr^{-1} in 1990. (We used a piecewise linear fit to the data through zero in 1861, 35 Tg S yr^{-1} in 1953 and 68 Tg S yr^{-1} in 1973. The implied lower trend over 1973–90 is uncertain, but in accord with data for Europe⁴⁷ and with the recent reduction in emissions growth in the USA.) The lower panels show both modelled and observed temperatures, the latter from IPCC92³⁹ smoothed with a 10-year gaussian filter. Model projections use best-guess parameters.

For comparison, the 2100 values for the IPCC90 scenarios are also shown in Fig. 2.

The most important points to note are the substantially reduced forcing in each best-guess case relative to the corresponding base case; the wide range of forcing possibilities, all corresponding to the same basic 'existing policies' assumption; and the reduced range of uncertainty in the best guess compared with the base case, largely because the high forcing cases have their larger CO₂ effects offset more by aerosol forcing. The reduced forcing compared with what would have been obtained using the IPCC90 methodology arises primarily from our inclusion of CO₂ fertilization feedback and sulphate aerosol forcing. As can be seen from Table 1, the aerosol effect depends on the scenario: IS92b is similar to IS92a, scenarios e and f have larger negative aerosol forcing contributions, whereas c and d have small positive contributions. For the best-guess case, total forcing under the c and d scenarios is below the (policy-driven) IPCC90 B scenario. None of the IS92 scenarios, however, approaches the very low forcing projections of IPCC90 scenarios C and D.

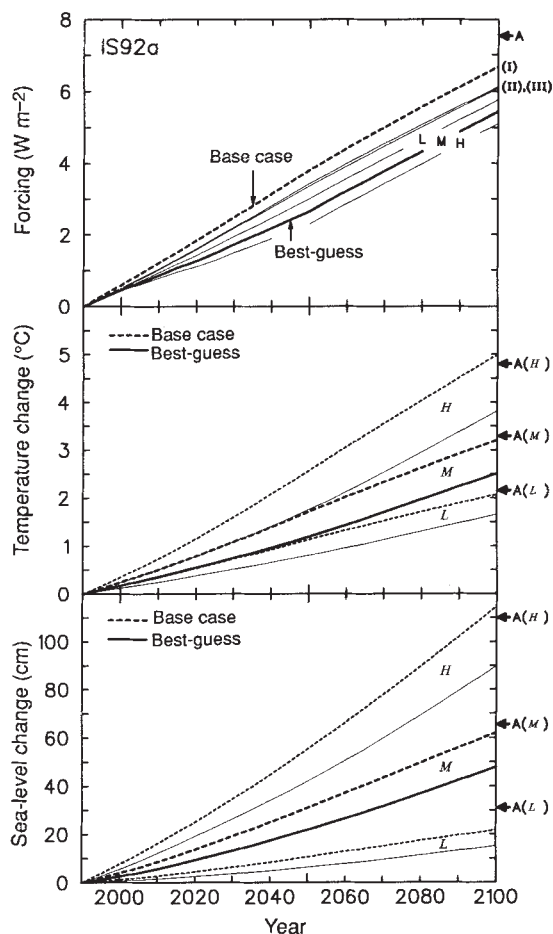


FIG. 4 Radiative forcing, global-mean temperature and global-mean sea level projections for emissions scenario IS92a. For forcing, the base case follows IPCC90 methodology (curve I). Curve II includes CO₂ fertilization feedback in calculating concentration trends. Curve III, slightly below II, additionally accounts for ozone-depletion feedback. L, M and H refer to low, middle and high aerosol forcing possibilities. Temperature and sea-level results are shown only for the base (dashed curves) and best-guess case forcings (full curves). Low, middle and high estimates are given (denoted L, M, H). For temperature, these are obtained using $\Delta T_{2x} = 1.5, 2.5$ and 4.5 °C with other model parameters kept at their best-guess values. For sea level, corresponding thermal expansion contributions are combined with ice-melt values based on low, middle and high model parameter values. For comparison, the equivalent 2100 low, middle and high estimates of temperature and sea level change for the IPCC90 Business as Usual scenario (SA90) are shown in the right (denoted A(L), A(M) and A(H)), with SA90 forcing shown by A in the top panel.

Temperature changes to 1990

Temperature projections for the six IS92 scenarios were made using our upwelling-diffusion energy-balance climate model³⁴, as used in IPCC90. Before describing the projections beyond 1990, however, we consider changes to 1990, as these affect the range of possible sulphate aerosol forcings and have implications for the climate sensitivity.

Climate models of the type used here incorporate very simplified representations of the physical processes involved. Nevertheless, they seem to simulate well the responses to external forcing produced by more complex models. They account for the lag effect of oceanic thermal inertia, and they allow one to assess the sensitivity of the results to model uncertainties, in particular to uncertainties in the climate sensitivity. The present model additionally differentiates between the hemispheres and so permits the consideration of forcing changes specific to one hemisphere, as is necessary if we are to account for the effects of sulphate aerosol forcing.

The most important model parameter is the climate sensitivity which is specified by the equilibrium global-mean temperature change for a CO₂ doubling. Largely on the basis of general circulation model experiments, this temperature change, ΔT_{2x} , is thought to lie in the range 1.5 – 4.5 °C, with a 'best-guess' value of 2.5 °C (ref. 1). The other parameters affecting the model's response are the mixed-layer depth (h), the oceanic vertical diffusivity (K), the upwelling rate (w) and the temperature change of high-latitude sinking water relative to the global-mean change (π). Sensitivity to these parameters is less than to ΔT_{2x} (ref. 35). We assume $h = 90$ m, $K = 1$ cm² s⁻¹, $w = 4$ m yr⁻¹ and $\pi = 0.2$. Slightly different values were used in IPCC90. The set of values used here is justified more fully in ref. 36. The greatest difference from IPCC is in our choice of π (IPCC90³⁷ used $\pi = 1$). Our choice of a lower value is supported by Schlesinger and Jiang³⁸. Smaller π produces larger surface warming and smaller oceanic thermal expansion. The net effect on sea level change is a small reduction.

Figure 3 shows the radiative forcing to 1990 and the corresponding modelled temperature changes, both for the global mean and for the Northern Hemisphere. Modelled temperature changes are shown only for $\Delta T_{2x} = 2.5$ °C. For comparison, observed temperature changes, as recently updated by IPCC³⁹, are also shown.

If ozone-depletion feedback and best-guess sulphate aerosol effects are included, then the total forcing change to 1990 is reduced markedly (by ~ 1 W m⁻² or 40%) from the IPCC90 estimate¹⁴. This brings the predictions of temperature change using $\Delta T_{2x} = 2.5$ °C into closer agreement with observations. In the absence of these two factors, the modelled changes with $\Delta T_{2x} = 2.5$ °C are appreciably larger than the observed changes; the value of ΔT_{2x} giving the best fit to the 1880–1990 warming is only 1.5 °C (compared with 1.4 °C in ref. 36, which used slightly different temperature data). For the case with ozone-depletion feedback and best-guess aerosol forcing, the modelled changes using $\Delta T_{2x} = 2.5$ °C are noticeably less than the observed changes. The required sensitivity to obtain the best fit to the observations is $\Delta T_{2x} = 3.4$ °C. If only best-guess aerosol forcing is included (without ozone-depletion feedback), our best-guess empirical value for the sensitivity is 3.0 °C (compare with ref. 40 for similar assumed forcing). The inclusion of the aerosol forcing flattens the model-projected warming trend over 1950–70, especially in the Northern Hemisphere, in qualitative agreement with observations.

Figure 3 also indicates the possible range of values for the negative aerosol forcing. Our largest negative values, which are at the smaller end of the range given by Charlson *et al.*¹⁰, lead to an overall negative forcing trend over 1860–1980 in the Northern Hemisphere. As a consequence, the modelled temperature change for the Northern Hemisphere over this period is much less than observed, and, more important, there is a marked difference in warming between the Northern and

Southern hemispheres which is not evident in the observations. It is possible that differential warming has been masked by natural variability, but this seems unlikely. On empirical grounds, therefore, and in accord with refs 29 and 40 and three-dimensional modelling results⁸, our choice for the range of likely aerosol forcing is more defensible than the values suggested in ref. 10 which were based on a simpler one-dimensional analysis.

Future temperature and sea level

Projections of temperature change are obtained by running the climate model with the various past forcing possibilities combined with consistent projected forcings to 2100. The number of model runs to be made is determined by the product of the number of future emissions scenarios (6), the number of forcing projections for each scenario (12, see below) and the number of sets of model parameters (3, see below). For each scenario, we have considered forcing projections with and without CO₂ fertilization feedback, with and without ozone-depletion feedback, and for low, middle and high aerosol forcing. This set of 12 possibilities gives an idea of the uncertainties in future forcing under any particular emissions scenario, although it is not a comprehensive analysis of uncertainty. For the climate model, we have carried out simulations for different climate sensitivity values, $\Delta T_{2x} = 1.5, 2.5$ and 4.5°C . As noted above, the sensitivity of the results to other model parameters is less than to ΔT_{2x} (see ref. 35) so these have been kept at their best-guess values. Only a selection of results is presented here.

The climate model also produces results for global-mean thermal expansion of the oceans, a principal component of future sea-level rise. To obtain total sea-level rise, meltwater contributions are added from separate, highly aggregated models representing small glaciers, the Greenland ice sheet and Antarctica, each driven by the global-mean temperature projection. The ice-melt models are similar to those used by IPCC90 (which we developed in conjunction with Warrick and Oerlemans⁴¹), with improvements as described in refs 35 and 42. Ice-melt uncertainties are accounted for by using a range of values for ice-melt model parameters^{35,42}.

Detailed results for IS92a are given in Fig. 4. This is the scenario most nearly equivalent to SA90⁴. The upper panel shows the sequential reductions in forcing that arise from including a CO₂ fertilization effect (I to II), accounting for ozone-depletion feedback (II to III) and accounting for the negative forcing due to sulphate aerosols (III to L, M or H). Including CO₂ fertilization to balance the contemporary carbon budget reduces the 1990–2100 forcing change by ~11% (compared with

a maximum of 22% for IS92c and a minimum of 7% for IS92e). The effect of including stratospheric-ozone-depletion feedback on post-1990 forcing is negligible (in all scenarios), whereas aerosols reduce the IS92a forcing by a further 6–19% relative to the base case. Overall, the best-guess forcing for IS92a is ~19% below the base case in the year 2100 (see Table 1 for results for other scenarios).

Low, middle and high temperature and sea-level projections for IS92a are shown in the lower panels of Fig. 4. For temperature, the warming over 1990–2100 is between 1.7°C and 3.8°C for the best-guess forcing case (compared with a warming range of 2.1°C to 5.0°C for base-case forcing). The corresponding sea-level rise lies between 15 and 90 cm for best-guess forcing and between 22 and 115 cm for base-case forcing. Temperature and sea-level change projections based on best-guess forcing are 20–24% and 21–30%, respectively, below those for base-case forcing.

The most direct comparison that can be made with IPCC90 is between IS92a and the SA90 results. The latter results are shown (for the year 2100 only) on the right-hand side of Fig. 4. IS92a and SA90 are comparable only in a limited sense, and differences arise because of differing emissions projections, revised methods for calculating the implied concentration and radiative forcing changes, changes in the assumed best-guess climate model parameters, and improvements in the ice-melt models. For temperature changes over 1990–2100, the IPCC90 best guess for SA90 (based on $\Delta T_{2x} = 2.5^\circ\text{C}$) was 3.3°C . Our corresponding best guess for IS92a is 2.5°C . For sea level, the equivalent projected changes are: SA90, 66 cm; IS92a, 48 cm. These are substantial reductions, but the projected changes in global-mean temperature and sea level are still very large. For comparison, the warming corresponds to a rate roughly five times that observed over the past century, and the sea level rise is at a rate roughly four times that estimated for the past century. The uncertainty ranges for these projections are large, and they are higher in the present results than in IPCC90.

Figure 5 summarizes temperature and sea-level projections for all scenarios for the overall best-guess cases ($\Delta T_{2x} = 2.5^\circ\text{C}$, best-guess ice-melt model parameters, and best-guess forcing). IPCC90 best-guess cases are also shown (2100 values). For the IS92 scenarios, the differences in temperature and sea level projections up to 2050 are fairly small, less than the forcing differences between scenarios (see Fig. 2). For the next 60 years or so, future temperature and sea-level changes are, therefore, relatively insensitive to the uncertainties in future emissions. This means that the IS92a results given in Fig. 4 can be taken as representative of all the emissions scenarios until about 2050.

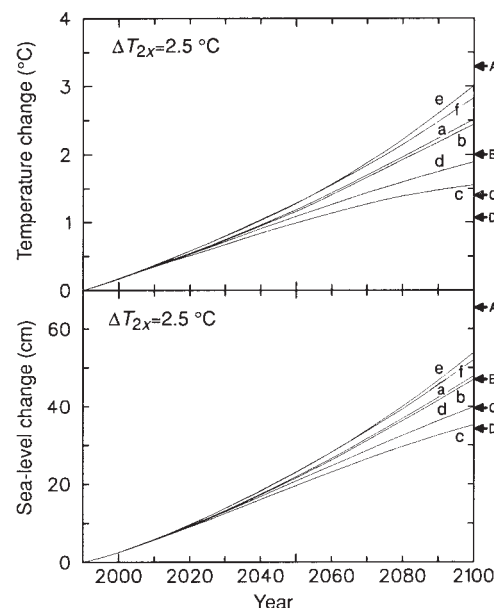


FIG. 5 Temperature and sea-level projections for the various IS92 emissions scenarios for the best-guess sets of climate and ice-melt model parameters (including $\Delta T_{2x} = 2.5^\circ\text{C}$). Note the similarity up to around 2050. Corresponding results for 2100 for the four emissions scenarios considered by IPCC90 are included on the right (A, B, C, D).

These projections are subject to other large uncertainties, particularly uncertainties in the climate sensitivity (Fig. 4). Reducing the error bounds on the climate sensitivity should therefore have high priority in greenhouse research^{36,43}.

Conclusions

These concentration, radiative forcing, temperature and sea-level projections are meant to complement the scientific update² of the 1990 IPCC report¹. We have addressed the new scientific issues raised in this update, the effects of CO₂ fertilization on CO₂ concentration projections, the possible negative feedback on halocarbon radiative forcing due to stratospheric ozone depletion, and the negative forcing due to sulphate aerosols. Largely because of these factors (specifically, the first and third), future projections of global-mean temperature and sea level change are reduced markedly (by 20–30%) below the equivalent projections made using the IPCC90 methodology (compare, for example, base-case and best-guess estimates in Fig. 4). But our projections, by including the effects of sulphate aerosols, introduce an additional complication: a reversed and enhanced differential in the rates of warming between the hemispheres. This could have an important influence on the climate system beyond the direct greenhouse effects currently considered by general circulation climate models.

The following important results have emerged. First, the effects so far of ozone-depletion feedback and sulphate aerosol forcing bring the best-guess empirical estimate of the climate sensitivity ($\Delta T_{2x} = 3.4^\circ\text{C}$) into closer agreement with the range of estimates based on general circulation climate model results. If both of these factors are ignored, the best-guess empirical sensitivity is only 1.5°C . The improved agreement here is encouraging, but it should not be over-interpreted. If other uncertainties are taken into account³⁶, then one finds a very wide range of ΔT_{2x} values to be compatible with observed temperature changes.

This point reinforces our second result, namely that uncertainty in future greenhouse-gas-related temperature and sea-level changes is strongly affected by the climate sensitivity. Reducing uncertainties in this parameter should have high

priority, although the best strategy for this is far from clear. The main reason for our conclusion here is that projections of temperature and sea-level changes out to 2050 have only a small range for any given ΔT_{2x} value, but differ widely according to the chosen value of ΔT_{2x} . The similarity between scenarios should not be interpreted as implying a general insensitivity to future emissions trajectories. Rather, it emphasizes the very long-term nature of the greenhouse problem. A more salutary result is the fact that, beyond the middle of next century, projections for the various scenarios diverge rapidly. Inertia in population growth and socio-economic systems would make it difficult to shift from one scenario to another.

Climate sensitivity is not the only uncertainty that needs to be reduced. In the scenarios in which SO₂ emissions increase markedly (a, b, e and f), an important additional uncertainty arises through not being able to quantify the effect of aerosols to better than around $\pm 50\%$. In scenario IS92a, for example, this translates to an uncertainty of $\pm 6\%$ in total forcing, $\pm 8\%$ in temperature change and $\pm 9\%$ in sea-level change (over 1990–2100). Uncertainties arising from our inadequate understanding of the global carbon cycle, judged from a comparison of results with and without CO₂ fertilization feedback, are of comparable magnitude.

The projections presented here show only the anthropogenic component of future change, and natural variability will be superimposed on this. Natural variability includes both internally generated variability⁴⁴ and the effects of external forcing agents⁴⁵, the latter including short-term factors like the recent volcanic eruption of Mount Pinatubo. The expected anthropogenic rates of change, however, are much greater than the corresponding natural rates of change over intervals of 30 years or longer, becoming increasingly so as the interval increases. The reduced rates of warming and sea-level rise that we have obtained compared with IPCC90 are still four to five times those that have occurred over the past century. Such changes are certain to present a considerable challenge for humanity. □

T. M. L. Wigley and S. C. B. Raper are at the Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

- Houghton, J. T., Jenkins, G. J. & Ephraums, J. J. (eds) *Climate Change. The IPCC Scientific Assessment* (Cambridge Univ. Press, 1990).
- Houghton, J. T., Callander, B. A. & Varney, S. K. (eds) *Climate Change 1992. The Supplementary Report to the IPCC Scientific Assessment* (Cambridge Univ. Press, 1992).
- Pepper, W. J. et al. *Emission Scenarios for the IPCC—an update: Background Documentation on Assumptions, Methodology, and Results*. (US Environmental Protection Agency, Washington DC, 1992).
- Leggett, J. A., Pepper, W. J. & Swart, R. J. in *Climate Change 1992. The Supplementary Report to the IPCC Scientific Assessment* (eds Houghton, J. T., Callander, B. A. & Varney, S. K.) 69–95 (Cambridge Univ. Press, 1992).
- Mitchell, J. F. B. & Gregory, J. M. in *Climate Change 1992. The Supplementary Report to the IPCC Scientific Assessment* (eds Houghton, J. T., Callander, B. A. & Varney, S. K.) 171–175 (Cambridge Univ. Press, 1992).
- Ramaswamy, V., Schwarzkopf, M. D. & Shine, K. P. *Nature* **355**, 810–812 (1992).
- Isaksen, I., Ramaswamy, V., Rodhe, H. & Wigley, T. M. L. in *Climate Change 1992. The Supplementary Report to the IPCC Scientific Assessment* (eds Houghton, J. T., Callander, B. A. & Varney, S. K.) 47–67 (Cambridge Univ. Press, 1992).
- Charlson, R. J., Langner, J., Rodhe, H., Leovy, C. B. & Warren, S. G. *Tellus AB* **43**, 152–163 (1991).
- Wigley, T. M. L. *Nature* **349**, 503–506 (1991).
- Charlson, R. J. et al. *Science* **255**, 423–430 (1992).
- Nordhaus, W. D. & Yohe, G. W. in *Changing Climate* 87–153 (National Academy Press, Washington DC, 1983).
- Edmonds, J. A., Reilly, J. M., Gardner, R. H. & Brenkert, A. *Uncertainty in Future Global Energy Use and Fossil Fuel CO₂ Emissions 1975 to 2075*, TR036 (US Dept of Energy, Carbon Dioxide Research Division, Washington DC, 1986).
- Wigley, T. M. L., Holt, T. & Raper, S. C. B. *STUGE (an Interactive Greenhouse Model): User's manual* (Climatic Research Unit, Norwich, 1991).
- Shine, K. P., Derwent, R. G., Wuebbles, D. J. & Morcrette, J.-J. in *Climate Change, The IPCC Scientific Assessment* (eds Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) 41–68 (Cambridge Univ. Press, 1990).
- Harvey, L. D. D. *Clim. Change* **15**, 343–381 (1989).
- Wigley, T. M. L. *Global biogeochem. Cycles* **5**, 373–382 (1991).
- Maler-Reimer, E. & Hasselmann, K. *Clim. Dynam.* **2**, 63–90 (1987).
- Wigley, T. M. L. *Tellus* (in the press).
- Watson, R. T., Rodhe, H., Oeschger, H. & Siegenthaler, U. in *Climate Change, The IPCC Scientific Assessment* (eds Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) 1–40 (Cambridge Univ. Press, 1990).
- Watson, R. T., Meira Filho, L. G., Sanhueza, E. & Janetos, A. in *Climate Change 1992. The Supplementary Report to the IPCC Scientific Assessment* (eds Houghton, J. T., Callander, B. A. & Varney, S. K.) 25–46 (Cambridge Univ. Press, 1992).
- Enting, I. & Mansbridge, J. *Tellus B43*, 156–170 (1991).
- Born, M., Door, H. & Levin, I. *Tellus B42*, 1–8 (1990).
- Osborn, T. J. & Wigley, T. M. L. *Clim. Dynam.* (submitted).
- Watson, R. T. & Albritton, D. L. *Scientific Assessment of Ozone Depletion: 1991* (World Meteorological Organization, Geneva, in the press).
- Vaghjiani, G. & Ravishankara, A. *Nature* **350**, 406–409 (1991).
- Lelieveld, J. & Crutzen, P. *Nature* **355**, 339–342 (1992).
- Lal, M. & Holt, T. *Clim. Res.* **1**, 85–95 (1991).
- Ramanathan, V., Cicerone, R. J., Singh, H. B. & Kiehl, J. T. *J. geophys. Res.* **90**, 5547–5566 (1985).
- Wigley, T. M. L. *Nature* **339**, 356–367 (1989).
- Charlson, R. J., Langner, J. & Rodhe, H. *Nature* **348**, 22 (1990).
- Twomey, S. A., Peipgrass, M. & Wolfe, T. L. *Tellus B* **36**, 356–366 (1984).
- Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. *Nature* **326**, 655–661 (1987).
- Albrecht, B. A. *Science* **245**, 1227–1230 (1989).
- Wigley, T. M. L. & Raper, S. C. B. *Nature* **330**, 127–131 (1987).
- Wigley, T. M. L. & Raper, S. C. B. in *Climate and Sea Level Change: Observations, Projections and Implications* (eds Warrick, R. A., Barrow, E. M. & Wigley, T. M. L.) (Cambridge Univ. Press, in the press).
- Wigley, T. M. L. & Raper, S. C. B. in *Climate Change: Science, Impacts and Policy* (eds Jäger, J. & Ferguson, H. L.) 231–242 (Cambridge Univ. Press, 1991).
- Bretherton, F. P., Bryan, K. & Woods, J. D. in *Climate Change, The IPCC Scientific Assessment* (eds Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) 173–193 (Cambridge Univ. Press, 1990).
- Schlesinger, M. E. & Jiang, X. *Nature* **350**, 219–221 (1991).
- Folland, C. K. et al. in *Climate Change 1992. The Supplementary Report to the IPCC Scientific Assessment* (eds Houghton, J. T., Callander, B. A. & Varney, S. K.) 135–170 (Cambridge Univ. Press, 1992).
- Schlesinger, M. E., Jiang, X. & Charlson, R. J. in *Proc. Conf. on Global Change: its Mitigation through Improved Production and Use of Energy* (eds Allred, J. C. & Nichols, A. S.) (American Institute of Physics, in the press).
- Warrick, R. A. & Oerlemans, H. in *Climate Change, The IPCC Scientific Assessment* (eds Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) 257–281 (Cambridge Univ. Press, 1990).
- Raper, S. C. B., Wigley, T. M. L. & Warrick, R. A. in *Proc. SCOPE Workshop on Rising Sea Level and Subsiding Coastal Areas* (ed. Milliman, J. D.) (Wiley, Chichester, in the press).
- Hansen, J. E., Rossow, W. & Fung, I. *Issues Sci. Technol.* **7**, 62–69 (1990).
- Wigley, T. M. L. & Raper, S. C. B. *Nature* **344**, 324–327 (1990).
- Hansen, J. E. & Lacis, A. A. *Nature* **346**, 713–719 (1990).
- Möller, D. *Atmos. Envir.* **18**, 19–27 (1984).
- Smith, F. B. *Q. J. R. met. Soc.* **117**, 657–683 (1991).

ACKNOWLEDGEMENTS. This work was supported by grants from the US Department of Energy, Atmospheric and Climate Research Division, the Commission of the European Communities (DGXII), and the UK Department of the Environment. Comments by K. P. Shine and P. M. Kelly helped to improve the manuscript. Computer-generated diagrams were produced by M. Salmon.

Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*

Frederick S. Gimble & Jeremy Thorner*

Division of Biochemistry and Molecular Biology, Department of Molecular and Cell Biology, Room 401 Barker Hall, University of California, Berkeley, California 94720, USA

An unusual protein splicing reaction joins the N-terminal segment (A) and the C-terminal segment (C) of the 119K primary translation product (ABC) of the yeast *VMA1* gene to yield a 69K vacuolar H⁺-ATPase subunit (AC) and an internal 50K polypeptide (B). This 50K protein is a site-specific DNA endonuclease that shares 34% identity with the homothallic switching endonuclease. The site cleaved by the *VMA1*-derived endonuclease exists in a *VMA1* allele that lacks the derived endonuclease segment of the open reading frame. Cleavage at this site only occurs during meiosis and initiates 'homing', a genetic event that converts a *VMA1* allele lacking the endonuclease coding sequence into one that contains it.

ENDONUCLEASES play a central part in genomic rearrangements in the cell. The endonuclease encoded by the homothallic switching (*HO*) gene¹ on chromosome IV of the yeast *Saccharomyces cerevisiae* initiates 'mating type switching' by cleaving at a site in the *MAT* locus on chromosome III (ref. 2). The double-strand break introduced by the *HO* enzyme is followed by a unidirectional gene conversion event that replaces the information at the *MAT* locus with information copied from either of two homologous loci that reside near the telomeres of chromosome III. A double-strand break followed by gap repair has even been proposed as a model for general recombination³.

Another genomic rearrangement that depends on the action of site-specific DNA endonucleases is the process of intron homing (for reviews, see refs 4, 5). The DNA endonucleases that initiate these events share certain amino-acid sequence motifs⁶⁻⁸ with *HO*, but are encoded by open reading frames in the middle of group I introns. At the molecular level, intron homing is similar to *HO*-mediated *MAT* switching; the intron-encoded endonuclease introduces a staggered double-strand break at a defined site in the recipient (intronless) allele which is then repaired with the information from the donor (intron-containing) allele. The first of this kind of mobile element to be identified was ω , an intron found in the gene for the 21S ribosomal RNA in the mitochondrial DNA of some strains of *S. cerevisiae*⁹. Subsequently, similar mobile introns were found in the mitochondria and chloroplasts of various eukaryotes¹⁰⁻¹², and even in bacteriophage T4 (ref. 5). We show here that such mobile DNA endonucleases are not confined to those encoded in introns.

VDE-mediated cleavage of DNA

We originally purified the 50K *VMA1*-derived endonuclease (VDE; relative molecular mass, 50,000) and isolated the gene encoding it because it was a novel factor that bound to the pheromone response element^{13,14}, a *cis*-acting site involved in controlling genes of the mating pathway in *S. cerevisiae* (F.S.G. and J.T., manuscript in preparation). Surprisingly, the gene we identified had been isolated previously, first as the *TFPI* gene¹⁵ (defined initially by a dominant mutation conferring resistance to the drug trifluoperazine), then later as the *VMA1* gene because it produces the 69K subunit of the vacuolar membrane H⁺-ATPase^{16,17}. The 69K ATPase subunit is generated by joining

the N- and C-terminal portions of the 119K primary translation product of the *VMA1* gene in a protein splicing reaction that liberates the internal 50K segment, which was termed the spacer protein¹⁶. We not only purified this 50K protein on the basis of its ability to bind to DNA, but also found that its deduced amino-acid sequence is 34% identical to that of the yeast *HO* endonuclease, as recognized previously¹⁷. To characterize the 50K protein, we purified it from yeast and from *Escherichia coli* cells carrying a plasmid expressing the *HO*-like coding region of *VMA1* at high levels.

We initially tested for endonuclease activity by incubating the purified recombinant 50K protein with DNA derived from bacteriophage λ and found that the 50K protein cleaved at a single site in the phage λ genome (Fig. 1a). Cleavage of λ DNA by VDE occurred at a much higher rate when Mn²⁺ was used as the cofactor instead of Mg²⁺. Mn²⁺ as a cofactor relaxes the specificity of type II restriction endonucleases, like *EcoRI* (ref. 18), and, more significantly, also reduces the specificity of *HO* endonuclease¹⁹. It is likely that Mn²⁺ has the same effect on VDE activity because, in the presence of Mn²⁺, several other sites in λ DNA were cleaved inefficiently when large amounts of VDE protein were used (data not shown).

By analogy with the intron-encoded endonucleases, we reasoned that VDE might cleave an 'empty' allele of the *VMA1* gene which was precisely deleted for the portion of the coding sequence that specifies the VDE protein (Fig. 1c). Indeed, the deletion allele (termed *VMA1 Δ vde*) was cleaved efficiently at a specific site by the VDE protein; in contrast, an intact *VMA1* gene was not cleaved (Fig. 1a). Moreover, the *VMA1 Δ vde* allele was readily cleaved even when Mg²⁺ was used as the cofactor. Mg²⁺ is likely to be the more physiologically relevant divalent cation because the intracellular concentration of Mg²⁺ is much higher than that of Mn²⁺ in yeast cells²⁰. Recombinant VDE protein purified from *E. coli* cleaved *VMA1 Δ vde* DNA as efficiently as purified VDE protein generated *in vivo* by protein splicing (Fig. 1b).

To determine the precise position of cleavage, we used primer elongation to map the VDE-catalysed cleavage site in λ DNA and in *VMA1 Δ vde* DNA. VDE-dependent cleavage of the *VMA1 Δ vde* allele occurred at a unique position in each strand of the DNA (Fig. 2a); results were similar for λ DNA (data not shown). In both substrates, VDE protein made a staggered break (see arrowheads and dashed line in Fig. 2b) and formed 3' overhangs of four bases. The cut DNA could be religated

* To whom correspondence should be addressed.

readily with T4 DNA ligase (data not shown), indicating that the ends generated were 5'-phosphoryl and 3'-hydroxyl. Intron-encoded endonucleases, I-SceI (ref. 21) and I-SceII (refs 12, 22), as well as HO (ref. 19) and Endo SceI²³, also leave 4-base 3' overhangs with 5'-phosphoryl and 3'-hydroxyl termini, even though each cleaves at a completely different nucleotide sequence^{12,21,22,24}.

The VDE cleavage sites in the λ and yeast substrates are substantially similar (Fig. 2b). Differences between the two substrates, such as the C·G→T·A substitution within the cleavage site itself, may explain why the λ substrate is only cleaved efficiently in the presence of Mn²⁺. The sequence of the pheromone response element, which we used to purify the VDE protein from yeast on the basis of its DNA-binding ability (F.S.G. and J.T., manuscript in preparation), is similar to that portion of the *VMA1Δvde* allele that is proximal to, but does not overlap, the cleavage site (Fig. 2b). Although DNA containing just a pheromone response element is not a substrate for cleavage, purified VDE protects such sequences from digestion by DNase I in footprinting experiments. These observations suggest that the recognition site for DNA-binding of VDE to the *VMA1Δvde* allele includes the pheromone response element-like region. The recognition sites for DNA-binding by the Endo SceI and I-SceII enzymes^{25,26} are a similar distance away from their cleavage sites.

Homing of the VDE element to *VMA1Δvde*

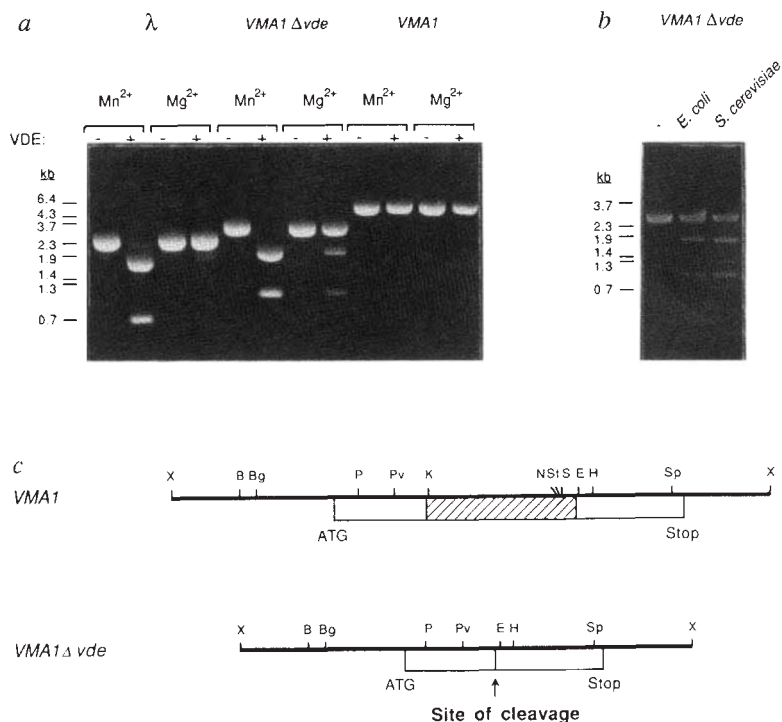
The site cleaved by VDE in an 'empty' *VMA1* gene is at exactly the position where the VDE coding segment is intercalated in a full-length *VMA1* gene (Figs 1c and 2b, caret symbol). This observation suggested that VDE might be capable of promoting homing of its own coding element *in vivo*. We therefore analysed DNA from the four haploid progeny of tetrads derived from sporulation of a *VMA1Δvde/VMA1* diploid cell. In this heterozygous diploid the wild-type *VMA1* allele (the donor) functions both as the source of VDE protein to effect cleavage and of VDE genetic information to use during gene conversion, whereas the *VMA1Δvde* allele (the recipient) functions as the target site. If no homing occurs, there should be strict mendelian segregation of the two alleles: two spores should contain *VMA1*

DNA and two spores should contain *VMA1Δvde* DNA. Strikingly, however, in 23 out of 26 tetrads analysed, the segregation pattern (*VMA1* : *VMA1Δvde*) was either 4:0 or 3:1, whereas only three displayed the 2:2 pattern diagnostic of mendelian segregation (Fig. 3). These results provided unequivocal evidence that gene conversion events were occurring at high frequency at this locus. Moreover, the gene conversion observed was always unidirectional, generating a larger number of donor alleles (*VMA1*) than recipient alleles (*VMA1Δvde*), as expected for a homing process.

To determine whether VDE protein was required for these events, a construction was made in which the VDE region of the *VMA1* gene was replaced by the *HIS3* gene (see Fig. 3 legend). This *vma1Δvde::HIS3* allele is incapable of expressing the VDE protein, but still contains the flanking DNA sequences necessary for gene conversion. The *vma1Δvde::HIS3* disruption also prevents production of functional vacuolar H⁺-ATPase, resulting in poor germination and slow growth on standard rich medium¹⁷, a readily identifiable phenotype. When a heterozygous *vma1Δvde::HIS3/VMA1Δvde* diploid was subjected to sporulation, poor growth: normal growth of haploid spores segregated 2:2 in all 52 tetrads examined (data not shown), suggesting that gene conversion cannot occur when VDE is absent. When a multicopy plasmid expressing the VDE protein was introduced into the *vma1Δvde::HIS3/VMA1Δvde* diploid, poor growth segregated 4:0 or 3:1 in 48 of 62 tetrads examined, consistent with a greatly enhanced frequency of meiotic gene conversion (data not shown). Thus, VDE protein is required to initiate the homing process. Furthermore, among the 4:0 and 3:1 tetrads, every viable slow-growing spore was His⁺ (data not shown), indicating that when VDE was supplied *in trans*, the *HIS3*-containing version of the *VMA1* gene was transferred. Therefore, sequences other than the VDE coding sequence itself can be directed to a *VMA1Δvde* allele.

The tetrads that displayed 3:1 segregation at *VMA1* argued that homing must have occurred during meiosis²⁷. That some tetrads displayed 4:0 segregation suggested that VDE might also promote mitotic gene conversion events. To investigate this possibility, the *VMA1Δvde/VMA1* diploid strain was plated, 25 individual colonies were picked, grown vegetatively for many

FIG. 1 Site-specific DNA cleavage by the VDE enzyme. **a**, DNA substrates (top line) were incubated in the absence (–) or presence (+) of purified recombinant VDE protein with either Mn²⁺ or Mg²⁺ and analysed by electrophoresis in an agarose gel. **b**, Comparison of the endonuclease activity of bacterially produced recombinant VDE (*E. coli*) and authentic VDE purified from yeast (*S. cerevisiae*). **c**, Schematic diagram and restriction map of the full-length *VMA1* gene and the *VMA1Δvde* allele. Open boxes correspond to the portions of the gene that encode the N- and C-terminal regions of the 69K subunit of the vacuolar H⁺-ATPase; the hatched box corresponds to the portion that encodes the VDE segment. Restriction enzymes: B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hind*III; K, *Kpn*I; N, *Nae*I; P, *Pst*I; Pv, *Pvu*II; S, *Sac*II; Sp, *Sph*I; St, *Stu*I; X, *Xba*I. **METHODS**. Natural yeast VDE and bacterially expressed VDE were purified as described (F.S.G. and J.T., manuscript in preparation). DNA substrates were: λ , a 2,096-bp *Bst*BI fragment of bacteriophage λ DNA (nucleotides 25,884–27,980); *VMA1Δvde*, a 2,606-bp *Bam*HI–*Sph*I fragment containing the *VMA1Δvde* deletion allele¹⁶; and *VMA1*, a 3,968-bp *Bam*HI–*Sph*I fragment of the intact *VMA1* locus¹⁶. DNA (~150 ng) was incubated for 2 h at 30 °C with either bacterially produced recombinant VDE or purified authentic yeast VDE (100 ng) in 50 mM KCl, 25 mM Tris-HCl, pH 7.5, 2.5 mM 2-mercaptoethanol, and either 2.5 mM MnCl₂ or MgCl₂ (for the λ DNA substrate), and in 150 mM KCl, 25 mM Tris-HCl, pH 8.5, 2.5 mM 2-mercaptoethanol, and either 2.5 mM MnCl₂ or MgCl₂ (for the yeast DNA substrates) in 15 μ l final volume, extracted with phenol, and electrophoresed in a 1% agarose gel which was subsequently stained with ethidium bromide.



generations, and DNA from these samples was analysed by quantitative hybridization analysis²⁸. No increase in the relative intensity of the *VMA1* band compared with the *VMA1Δvde* band was detected in any of the 25 samples (data not shown). Additionally, a plasmid carrying the *VMA1Δvde* allele was introduced into two unrelated *VMA1*⁺ haploid strains, YPH499 and EG123 (ref. 29). After many generations of growth, DNA was extracted from two independent cultures of each plasmid-containing haploid and the plasmids were recovered and propagated in *E. coli* for analysis. In all nine *E. coli* transformants examined, only a plasmid identical to the original *VMA1Δvde* plasmid was observed. These results suggest that if VDE can mediate homing in mitotic cells, it occurs at a much lower frequency than in meiotic cells.

To confirm that VDE-mediated cleavage occurs during meiosis, the *VMA1Δvde/VMA1* diploid was subjected to a synchronous sporulation regimen and its DNA analysed by hybridization analysis (Fig. 4). Two bands appeared on the autoradiogram during the time course that co-migrated with the products of VDE cleavage of *VMA1Δvde* DNA *in vitro*, and these species reached a maximum between 8 and 10 hours after the shift (Fig. 4a). The appearance and disappearance of these bands provided direct evidence for the presence *in vivo* of a transient double-strand break in the *VMA1Δvde* allele. A third band appeared in this experiment that was not observed in cleavage reactions *in vitro* and that hybridized only to a 5'-specific probe (data not shown). This third band may represent non-specific degradation of the larger (5') cleavage product, or a true intermediate in the gene conversion process. For example,

exonucleolytic resection of one of the *MAT* fragments generated by HO cleavage has been observed *in vivo*⁵⁰, and apparent intermediates that contain 3'-overhanging single-stranded tails have been detected during meiotic gene conversion at the yeast *ARG4* locus³¹.

The kinetic pattern (Fig. 4b) was fully consistent with a gene conversion mechanism. At time zero, there were equal amounts of both donor and recipient DNA and no intermediates were detected, further indicating that little or no homing occurs in mitotically growing cells. After a few hours lag, the amount of the recipient allele decreased and was mirrored by a corresponding increase in the amount of the donor allele, suggesting that the new donor alleles were created from the recipient alleles that were lost. As expected for intermediates, the amount of the cleavage products peaked when the fraction of the recipient allele that was cleaved reached its half-maximal level (Fig. 4b), and the amount of the cleavage products decreased to nearly undetectable at the time when no further change in the levels of the donor and recipient alleles was observed. Although the fraction of the recipient DNA that was cut by the end of the time course (20 h) was only 40% (Fig. 4b), it should be noted that the percentage of diploid cells in the population that actually sporulated was 53%, as judged by direct counting of asci under a microscope. Thus if it is assumed that cleavage only occurs in cells that undergo sporulation, then the amount of cleavage observed can account for the fraction of the population that underwent homing during sporulation (88%; Fig. 3). This calculation suggests further that a homing event occurs successfully every time a recipient allele is cleaved.

FIG. 2 Location of VDE cleavage sites at the nucleotide level. *a*, Analysis of the VDE cleavage site in *VMA1Δvde* DNA. An end-labelled double-stranded substrate was incubated in the absence (–) or presence (+) of VDE protein and the product (arrowhead) generated from each strand (left, coding; right, non-coding) was compared with a sequencing ladder. The bond cleaved in each strand (arrow) is indicated. *b*, Alignment of the nucleotide sequence surrounding the cleavage sites (arrowheads and dashed lines) in *VMA1Δvde* with the pheromone response element (PRE) sequence and with the sequence surrounding the cleavage sites in λ DNA. Identities are given as white-on-black letters. The point of insertion of the VDE coding segment within an intact *VMA1* gene (caret symbol) is indicated.

METHODS. Cleavage site position was determined as described in ref. 12. A 758-bp *Pst*I–*Hind*III fragment containing the yeast *VMA1Δvde* cleavage site was inserted into the corresponding sites in Bluescript SK⁺ and KS⁺ (Stratagene). The 2,096-bp *Bst*BI fragment of phage λ that contains the VDE cleavage site was inserted into the *Cl* site of Bluescript KS⁺ in either orientation. Single-stranded DNA templates for the yeast and λ sequences were prepared from these plasmids by standard methods³⁹. Each template (1 μ g) was used to generate labelled complementary single strands using 1 pmol of an appropriate oligonucleotide primer (5'-CATAAGTGTAACTTCC-3' and 5'-AAGCTTTGTGTCACCC-3' for the λ template, and 5'-ATTCCAGGTGCTTTTGG-3' and 5'-TAGTACGCTTCATAATT-3' for the yeast template) using Sequenase enzyme (US Biochem). Products were extended with either unlabelled dideoxynucleotides (to generate a sequencing ladder) or with unlabelled deoxynucleotides (1mM of each dNTP) to generate the end-labelled double-stranded substrate used in the cleavage reaction. Substrates were incubated as described in Fig. 1 legend, either alone or with purified bacterially produced VDE (30 ng), and analysed in a 6% polyacrylamide gel adjacent to sequencing ladders prepared from the same templates.

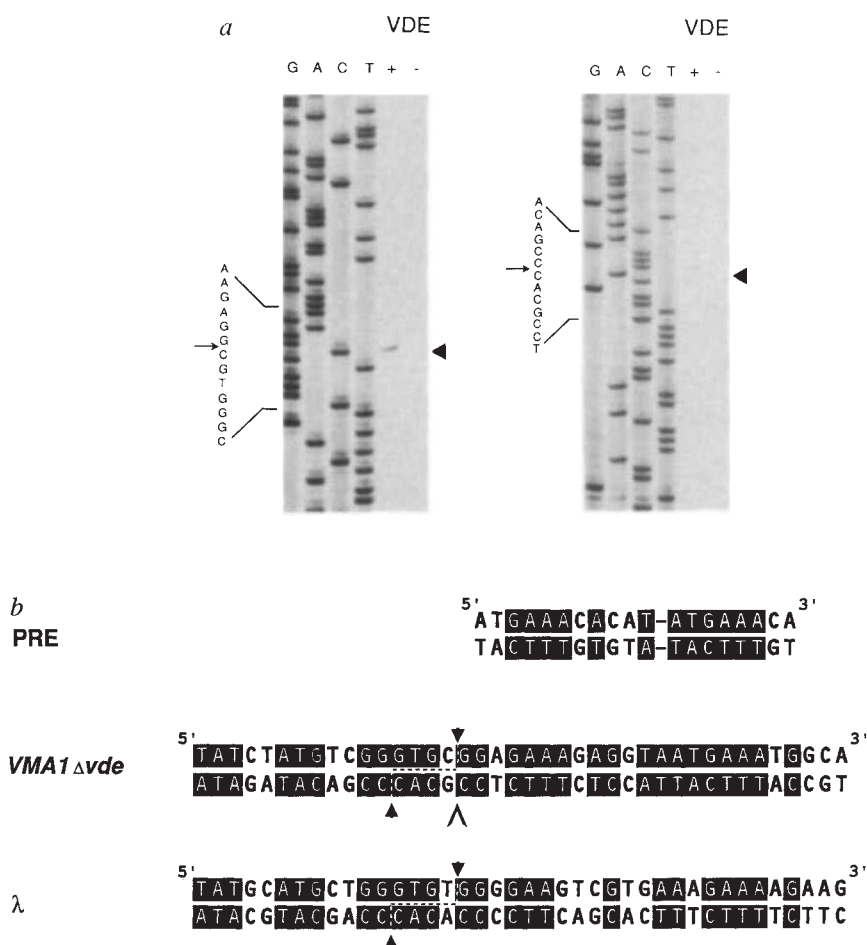
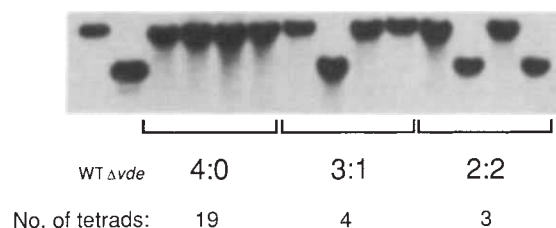


FIG. 3 Homing of the VDE coding segment by meiotic gene conversion. Hybridization analysis of genomic DNA extracted from spores of representative tetrads that segregated 4:0, 3:1 or 2:2 for the *VMA1* and the *VMA1Δvde* alleles, respectively. The number of tetrads displaying each segregation pattern is indicated.

METHODS. The *VMA1* locus in a *MAT a* strain (YPH499) (ref. 40) was transplaced⁴¹ with the *VMA1Δvde* deletion allele using linearized plasmid pVMAΔvde (3.5 kb *Bam*HI–*Sac*I fragment of ptfp-Δ10 (ref. 16) in pRS306 (ref. 40)). The resulting strain, YPH499-Δvde7D (Δvde), was mated with an isogenic *MATαVMA1*⁺ partner, YPH500 ('WT')⁴⁰, to yield a heterozygous diploid strain (YFG501). After sporulation⁴², DNA was prepared⁴³ from all four spores of 26 separate tetrads. Samples (~10 μg) from each were digested with *Bam*HI and *Sph*I and analysed by Southern blotting²⁸ using a radiolabelled probe specific for the *VMA1* locus (1,261-bp *Bgl*II–*Pvu*II fragment, see Fig. 1c). Transplacement of the VDE portion of the *VMA1* gene with the *HIS3* gene in strain YPH500 (ref. 40) was done⁴⁴ using a 4.5-kb fragment from plasmid pVMAΔvde-*HIS3*. Plasmid pVMAΔvde-*HIS3* was constructed by inserting the 5.5-kb *Xba*I fragment from pDW21¹⁶ containing *VMA1* into the *Xba*I site of pGEM-1 (Promega), and the VDE region (*Kpn*I–*Sac*I



fragment, see Fig. 1c) of the resulting plasmid was replaced with a *Bam*HI fragment containing the *HIS3* gene from pJ215 (ref. 45). The resulting *vma1Δvde::HIS3* strain (YFG502) was mated with strain YPH499-Δvde7D to generate a heterozygous *vma1Δvde::HIS3/VMA1Δvde* diploid strain (YFG503). Before sporulation of YFG503, it was transformed either with a plasmid (pADHVDE) that overexpresses VDE by ~10–20 fold (F.S.G. and J.T., manuscript in preparation) or with the empty control vector (pAD4M)⁴⁶.

A naturally occurring *VMA1* deletion allele

The role of I-SceI endonuclease in catalysing the rapid and irreversible dissemination of the ω intron into mitochondrial 21S rRNA genes in a population of yeast cells that lack the ω intron is well documented³². By analogy, the VDE element should also spread within a population by propagating itself into *VMA1Δvde* alleles as a result of VDE-mediated homing whenever diploid cells undergo sporulation. As this process had possibly not yet occurred in certain yeast stocks, we undertook a search for naturally occurring *VMA1* alleles apparently lacking a VDE segment. Genomic DNA of one wild wine yeast (DH1-1A) displayed a pattern of bands identical to that of the *VMA1Δvde/VMA1* diploid strain that we constructed in the laboratory (Fig. 5b; the upper band resulted from partial digestion). Initially, it was puzzling that this apparent deletion allele persisted in the same strain as an apparently full-length *VMA1* allele, because an 'empty' allele should be converted to the full-length allele during meiosis by VDE-mediated homing. One explanation for the co-existence of the two *VMA1* alleles was that this strain had never undergone a successful meiosis; indeed, we were unable to induce the DH1-1A strain to sporulate and form viable progeny.

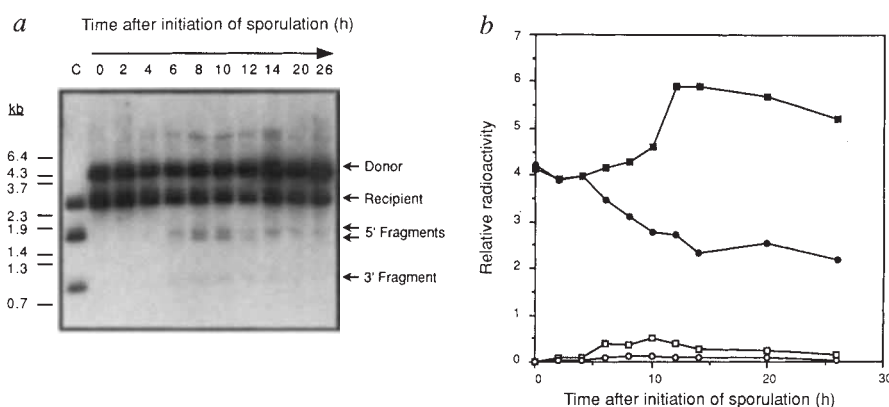
Alternatively, the shorter *VMA1* allele found in DH1-1A could lack a cleavage site for VDE, or the larger *VMA1* allele in DH1-1A might be unable to generate any functional VDE.

Extracts of the DH1-1A strain contain a 50K protein that cross-reacts on immunoblots with the anti-*S. cerevisiae* VDE antibodies raised previously, and this protein was purified by the same procedures as we used for VDE purification from our standard laboratory strains (Fig. 5a). The presence of this species suggests that protein splicing of the primary translation product of the larger *VMA1* allele occurs in the DH1-1A strain. As before, normal VDE protein purified from strain YPH499 cleaved *VMA1Δvde* DNA in the presence of either Mg^{2+} or Mn^{2+} , but the 50K protein from strain DH1-1A could cleave this substrate only in the presence of Mn^{2+} (Fig. 5b). In the reciprocal experiment, DNA from strain DH1-1A appeared to contain a VDE cleavage site because it was cut by both normal VDE protein and the 50K protein from DH1-1A, but only in the presence of Mn^{2+} . To detect digestion of the DH1-1A substrate in the presence of Mg^{2+} , a large amount of purified recombinant VDE protein had to be used (Fig. 5b). Thus in DH1-1A cells both the VDE endonuclease and its target substrate have diverged from their normal counterparts, perhaps explaining the co-existence of both a full-length *VMA1* gene and a recipient deletion allele in this strain.

Discussion

We have found that a 50K protein, whose function was not previously known, is a site-specific endonuclease (VDE) that is

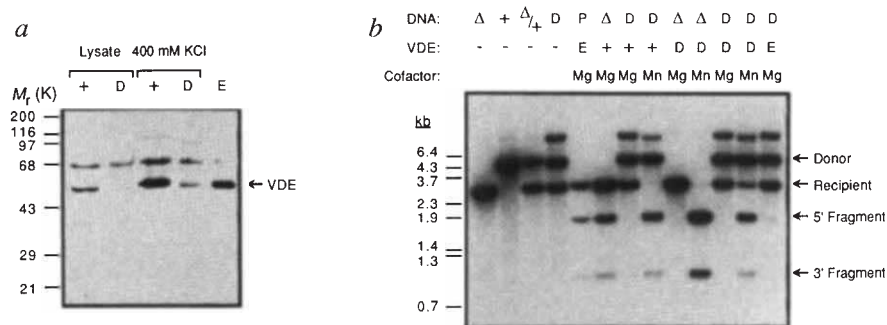
FIG. 4 Cleavage of the *VMA1Δvde* allele *in vivo* during sporulation. a, Hybridization analysis of genomic DNA extracted from samples, withdrawn at the indicated times, of a culture of a *VMA1Δvde/VMA1* diploid strain induced to undergo synchronous sporulation⁴⁷. Both donor and recipient alleles, as well as cleavage products, were detected by hybridization to a mixture of probes specific for the 5' and 3' regions of the *VMA1* coding sequence. Migration positions (arrows) of the bands representing the donor and recipient alleles, and the 5' and 3' fragments resulting from cleavage of the recipient *VMA1Δvde* allele are compared with the control fragments (C) generated by VDE-mediated cleavage *in vitro* of genomic DNA prepared from an *VMA1Δvde* haploid strain. b, The relative radioactivity, determined by normalizing to an internal standard in each lane, of the *VMA1* donor allele (filled squares), the *VMA1Δvde* recipient allele (filled circles), the sum of the two 5' products (squares), and the 3' product (circles), are plotted as a function of time after the initiation of sporulation.



METHODS. Samples (30 ml) of the culture (YFG501) were removed at various times following the shift to sporulation medium and DNA was prepared from each⁴⁸. DNA (~10 μg) from each time point was digested with *Bam*HI and *Sph*I and analysed by Southern blotting²⁸ using radiolabelled probes: a 1,261-bp *Bgl*II–*Pvu*II fragment (to detect the 5' portion of the *VMA1* gene);

a 852-bp *Eco*RI–*Sph*I fragment (to detect the 3' portion of the *VMA1* gene); and, when necessary, a 1,198-bp *Kpn*I–*Stu*I fragment (to detect the VDE segment of the *VMA1* gene) (Fig. 1c). Radioactivity in each band was determined by integration using a PhosphorImager (Molecular Dynamics) using ImageQuant software. Radioactivity in each band was normalized to that of an internal standard in each lane using an appropriate probe (*CMK1* gene⁴⁹). DNA from strain YFG499-Δvde7D was digested with bacterially produced VDE protein (150 ng) as described in Fig. 1 legend.

FIG. 5 VDE protein and VDE cleavage site in strain DH1-1A. *a*, Immunoblot of purified VDE (arrow) produced from *E. coli* (E) and VDE protein in crude lysates and in the eluate from Affi-gel blue chromatography of extracts from strain YPH499 (+) and from strain DH1-1A (D). *b*, Hybridization analysis of genomic DNAs incubated *in vitro* with either Mg^{2+} or Mn^{2+} in the absence (–) or presence of either purified bacterially produced VDE (E), or VDE purified from yeast strain YPH499 (+), or VDE purified from the DH1-1A strain (D). Substrates were: the *VMA1*Δ*vde* allele from YFG499-Δ*vde*7D (Δ); the full-length *VMA1* gene from YPH499 (+); both *VMA1*Δ*vde* and *VMA1* from YFG501 (Δ/+); DNA from the DH1-1A strain (D); plasmid p*VMA1*Δ*vde* (P). Migration positions (arrows) of the *VMA1* (donor) allele, the *VMA1*Δ*vde* (recipient) allele, the 5' cleavage product, and the 3' cleavage product are indicated.



METHODS. Bacterially expressed VDE was purified as described in Fig. 1 legend. To purify VDE from strains YPH499 and DH1-1A, frozen cells collected from 1 litre of culture in late exponential phase were washed in extraction buffer (0.2 M Tris-Cl, pH 8.0, 10 mM $MgCl_2$, 10% glycerol (v/v), 0.4 M $(NH_4)_2SO_4$, 1 mM EDTA) and lysed by vortexing with glass beads in a Bead-beater (Biospec). The lysate was clarified by centrifugation at low-speed and at 100,000g. Some clarified lysate was saved for immunoblot analysis and the remainder was fractionated using ammonium sulphate. The precipitated protein was resuspended in 25 mM HEPES, pH 7.6, 10% glycerol (v/v), 50 mM KCl and 0.1 mM EDTA containing 1 mM PMSF, dialysed against the same buffer, and loaded onto a bed of Affigel Blue (Bio-Rad) equilibrated in the same buffer. The column was washed with the same buffer and eluted

in two steps with the same buffer containing 150 mM and 400 mM (virtually all of the VDE was contained in this fraction). Eluates were concentrated 5-fold using Centricon-30 filtration units (Amicon), dialysed against the same buffer containing 50 mM KCl, and frozen. For immunoblots, recombinant VDE protein (75 ng), crude lysate (2.5 μ g) and Affigel Blue eluate (3 μ g) were electrophoresed on SDS-polyacrylamide gels⁵⁰, transferred to nitrocellulose and incubated with affinity-purified rabbit anti-VDE antibodies (F.S.G. and J.T., manuscript in preparation) and secondary antibodies coupled to horseradish peroxidase for chemiluminescent detection (ECL, Amersham). Genomic DNA was analysed as described in Fig. 4 legend. Cleavage of genomic DNA (10 μ g) by purified VDE (2.3 μ g Affi-gel blue eluates or 4.5 μ g bacterially produced enzyme) was as for Fig. 1 for yeast substrates, except that incubation was for 1 h. Plasmid DNA (2 ng) was digested under the same conditions with *E. coli*-produced VDE (150 ng).

generated when the 119K precursor protein encoded by the *VMA1* gene^{16,17} of *S. cerevisiae* undergoes an unusual auto-cleavage and religation reaction¹⁶. Our findings demonstrate that VDE initiates homing of its own coding sequence to *VMA1* alleles that do not contain it and represent the first example of homing for a gene segment that is not part of a group I intron. Intron homing events mediated by the I-SceI or I-SceII endonucleases occur immediately after cells containing the mobile element are mated to cells without it^{12,33}; HO-mediated MAT switching occurs only during mitotic S phase³⁴ and only in haploid cells (which are incapable of meiosis). In contrast, VDE catalyses its homing event only in meiotic cells. It is unclear, however, what restrains VDE action in mitotic cells, given that active VDE enzyme can be readily purified from vegetatively growing haploid or diploid cells. Perhaps VDE is capable of cleaving its target sequence in mitotic cells, but such breaks are repaired as rapidly as they are made. This seems unlikely because the breaks produced by HO cleavage both at MAT and at sites introduced artificially elsewhere in the genome persist for long enough to be detectable by our methods^{2,30}. Regulation of VDE activity could be achieved by several mechanisms: (1) VDE may only be allowed to enter the nucleus during meiosis; (2) an inhibitor that may be present only in mitotic cells blocks VDE action *in vivo*; (3) a cofactor present only during meiosis may be required for VDE function; and (4) some feature of chromosome structure or chromatin assembly specific to meiosis, such as formation of a synaptonemal complex, may be required for VDE-mediated cleavage and conversion. With regard to the last possibility, HO sites in the silent loci (*HMR* and *HML*) are normally inaccessible to cleavage unless chromatin structure is perturbed, for example in cells carrying *sir* mutations³⁵.

The existence of a naturally occurring deletion allele that lacks a VDE segment (Fig. 5b) suggests that VDE coding information was introduced relatively recently into the *VMA1* gene, and that the VDE segment was originally an independent mobile element. This conclusion is supported by the fact that the 69K vacuolar H^+ -ATPase subunits of most other fungal species are highly related in amino-acid sequence to the yeast protein, but are encoded by genes that do not contain any additional VDE-like sequence³⁶. It is interesting, however, that the *VMA1* homologue of the pathogenic yeast *Candida tropicalis* does contain a VDE-like segment in its coding sequence (G. E. Dean,

personal communication). A potential source of the VDE element may have been the *HO* gene (or its progenitor) because VDE and HO are homologous over almost their entire length and because *VMA1* and *HO* map within 20 centimorgans of one another on the left arm of chromosome IV (F.S.G. and J.T., manuscript in preparation).

Endonucleases integrated into group I introns are not deleterious to their host genes because they have found a 'safe haven' in self-splicing introns⁵. In the case of VDE, an endonuclease has become incorporated within a single continuous open reading frame. The host protein that has accepted incursion of the VDE segment is expressed in functional form not by elimination of an intron by splicing of messenger RNA, but rather by excision of the 50K VDE segment by splicing at the protein level¹⁶. How the VDE segment came to be tolerated as a molecular endosymbiont within the *VMA1* open reading frame is unclear. Perhaps VDE had inherent peptide hydrolase and transpeptidase activities in addition to its endonuclease activity before its integration into *VMA1*. There is evidence that both the I-SceII and the Endo SceI endonucleases may be derived from larger protein precursors^{37,38}, but it is not yet known whether protein splicing reactions or more typical proteolytic processing events are involved in maturation of these precursors. □

Received 20 February; accepted 9 April 1992.

- Kostriken, R., Strathern, J. N., Klar, A. J. S., Hicks, J. B. & Heffron, F. *Cell* **35**, 167–174 (1983).
- Herskowitz, I. & Jensen, R. E. *Meth. Enzym.* **194**, 132–146 (1991).
- Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. & Stahl, F. *Cell* **33**, 25–35 (1983).
- Perlman, P. S. & Butow, R. A. *Science* **246**, 1106–1109 (1989).
- Belfort, M. A. *Rev. Genet.* **24**, 363–385 (1990).
- Michel, F., Jacquier, A. & Dujon, B. *Biochimie* **64**, 867–881 (1982).
- Waring, R. B., Davies, R. W., Scacciochio, C. & Brown, T. A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6332–6336 (1982).
- Hensgens, L. A. M., Bonen, L., de Haan, M., van der Horst, G. & Grivell, L. *Cell* **32**, 379–389 (1983).
- Dujon, B. *Cell* **20**, 185–197 (1980).
- Lemieux, B., Turmel, M. & Lemieux, C. *Molec. gen. Genet.* **212**, 48–55 (1988).
- Muscarella, D. E. & Vogt, V. M. *Cell* **56**, 443–454 (1989).
- Wenzlau, J. M., Saidanha, R. J., Butow, R. A. & Perlman, P. S. *Cell* **56**, 421–430 (1989).
- Van Arsdell, S. W. & Thorner, J. in *Transcriptional Control Mechanisms* (eds Granner, D., Rosenfeld, M. G. & Chang, S.) 325–332 (Liss, New York, 1987).
- Kronstad, J. W., Holly, J. A. & MacKay, V. L. *Cell* **50**, 369–377 (1987).
- Shih, C.-K., Wagner, R., Feinstein, S., Kanik-Ennulat, C. & Neff, N. *Molec. cell. Biol.* **8**, 3094–3103 (1988).
- Kane, P. M. *et al. Science* **250**, 651–657 (1990).
- Hirata, R. *et al. J. biol. Chem.* **265**, 6726–6733 (1990).
- Hsu, M. & Berg, P. *Biochemistry* **17**, 131–138 (1978).

19. Kostriken, R. & Heffron, F. *Cold Spring Harb. Symp. quant. Biol.* **49**, 89–96 (1984).
20. Sarwar, G., Shah, B. G., Mongeau, R. & Hoppner, K. *J. Food Sci.* **50**, 353–357 (1985).
21. Colleaux, L., D'Auriol, L., Galibert, F. & Dujon, B. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6022–6026 (1988).
22. Delahodde, A. *et al. Cell* **56**, 431–441 (1989).
23. Watabe, H.-o., Iino, T., Kaneko, T., Shibata, T. & Ando, T. *J. biol. Chem.* **258**, 4663–4665 (1983).
24. Shibata, T., Watabe, H.-o., Kaneko, T., Iino, T. & Ando, T. *J. biol. Chem.* **259**, 10499–10506 (1984).
25. Kawasaki, K., Takahashi, M., Natori, M. & Shibata, T. *J. biol. Chem.* **266**, 5342–5347 (1991).
26. Wernette, C. *et al. Molec. cell. Biol.* **12**, 716–723 (1992).
27. Fogel, S., Mortimer, R. K. & Lusnak, K. in *The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance* (eds Strathern, J. N., Jones, E. W. & Broach, J. R.) 289–339 (Cold Spring Harbor Laboratory, New York, 1981).
28. Southern, E. *J. Molec. Biol.* **98**, 503–517 (1975).
29. Siliciano, P. G. & Tatchell, K. *Cell* **37**, 969–978 (1984).
30. White, C. I. & Haber, J. E. *EMBO J.* **9**, 663–673 (1990).
31. Sun, H., Treco, D. & Szostak, J. W. *Cell* **64**, 1155–1161 (1991).
32. Jacquier, A. & Dujon, B. *Cell* **41**, 383–394 (1985).
33. Zinn, A. R. & Butow, R. A. *Cell* **40**, 887–895 (1985).
34. Nasmyth, K. *Nature* **302**, 670–676 (1983).
35. Klar, A. J. S., Strathern, J. N. & Abraham, J. A. *Cold Spring Harb. Symp. quant. Biol.* **49**, 77–88 (1984).
36. Bowman, E. J., Tenney, K. T. & Bowman, B. J. *J. biol. Chem.* **263**, 13994–14001 (1988).
37. Nakagawa, K.-I., Hashikawa, J.-I., Makino, O., Ando, T. & Shibata, T. *Eur. J. Biochem.* **171**, 23–29 (1988).
38. Wernette, C. M., Saldanha, R., Perlman, P. S. & Butow, R. A. *J. biol. Chem.* **265**, 18976–18982 (1990).
39. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1989).
40. Sikorski, R. S. & Hieter, P. *Genetics* **122**, 19–27 (1989).
41. Scherer, S. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4951–4955 (1979).
42. Sherman, F., Fink, G. R. & Hicks, J. B. *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory, New York, 1986).
43. Hoffman, C. S. & Winston, F. *Gene* **57**, 267–272 (1987).
44. Rothstein, R. J. *Meth. Enzym.* **101**, 202–211 (1983).
45. Jones, J. S. & Prakash, L. *Yeast* **6**, 363–366 (1990).
46. Ballister, R. *et al. Cell* **59**, 681–686 (1989).
47. Kassir, Y. & Simchen, G. *Meth. Enzym.* **194**, 94–110 (1991).
48. Murray, J. A. H. & Cesareni, G. *EMBO J.* **5**, 3391–3399 (1986).
49. Pausch, M. H., Kaim, D., Kunisawa, R., Admon, A. & Thorner, J. *EMBO J.* **10**, 1511–1522 (1991).
50. Dreyfus, G., Adam, S. A. & Choi, D. Y. *Molec. cell. Biol.* **4**, 415–423 (1984).

ACKNOWLEDGEMENTS. We thank P. Kane and T. Stevens for plasmid DNA and for advice, I. Herskowitz for wine yeasts, R. Schekman for use of his PhosphorImager, and J. Trueheart and J. Rine for discussion. This work was supported by a postdoctoral fellowship from the Damon Runyon-Walter Winchell Cancer Research Fund (F.S.G.), by an NIH research grant (J.T.) and through facilities provided by the Cancer Research Laboratory of the University of California at Berkeley.

LETTERS TO NATURE

Pulsed high-energy γ -radiation from Geminga (1E0630+178)

D. L. Bertsch*, K. T. S. Brazier†, C. E. Fichtel*, R. C. Hartman*, S. D. Hunter*, G. Kanbach†, D. A. Kniffen‡, P. W. Kwok*, Y. C. Lin§, J. R. Mattox*¶, H. A. Mayer–Hasselwander†, C. v. Montigny†, P. F. Michelson§, P. L. Nolan§, K. Pinkau†, H. Rothermel†, E. J. Schneid||, M. Sommer†, P. Sreekumar* & D. J. Thompson*

* NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

† Max-Planck-Institut für Physik und Astrophysik,

Institut für Extraterrestrische Physik, D-8046 Garching, Germany

‡ Department of Physics, Hampden-Sydney College, Hampden-Sydney, Virginia 23943, USA

§ W. W. Hansen Experimental Physics Laboratory and Department of Physics, Stanford University, Stanford, California 94305, USA

|| Grumman Aerospace Corporation, Bethpage, New York 11714, USA

¶ CSC, Compton Observatory Science Support Center,

NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

HALPERN and Holt¹ have recently reported the detection of coherent pulsations with a period of 237 ms from the soft X-ray source 1E0630+178, which lies in the error box of the γ -ray source known as Geminga (2CG195+04). This observation provides compelling evidence that Geminga, an object whose nature has hitherto been mysterious, is an X-ray pulsar. Prompted by this discovery, we have searched the data from EGRET, the Energetic Gamma Ray Experiment Telescope on the Compton Gamma Ray Observatory, for a comparable signal in the γ -radiation from this part of the sky. We now report the detection of pulsed γ -rays, with energy >50 MeV, from 1E0630+178, confirming the identification of Geminga with this X-ray source. The period derivative, $(11.4 \pm 1.7) \times 10^{-15} \text{ s s}^{-1}$, suggests that Geminga is a nearby, isolated, rotating neutron star with a magnetic field of 1.6×10^{12} gauss, a characteristic age of $3 \times 10^5 \text{ yr}$ and a spin-down energy loss rate of $3.5 \times 10^{34} \text{ erg s}^{-1}$.

The γ -ray source Geminga (2CG195+04) was first seen by the SAS-2 satellite (refs 2–4) in observations made roughly 20 years ago. Then, as now, it was one of the three strongest high-energy γ -ray sources in the sky, the other two being the Crab and Vela pulsars PSR0531+21 and PSR0833+45. It was seen at roughly the same intensity both times that SAS-2 viewed the region of the sky where it was located, specifically 14–21

December 1972 and 17–24 April 1973. Subsequently, the COS-B satellite observed the source five times between 1975 and 1982 (refs 5–7), again with essentially the same intensity each time.

After the original SAS-2 observations, Thompson *et al.*⁴ argued on the basis of the strength, its location, the absence of observations at other wavelengths and other factors that it was highly probable that the source was galactic, and possible that it was an undiscovered pulsar. Subsequent attempts to find a radio counterpart failed even after the COS-B results provided a better location⁸. Other theories were put forward over the years, but the mystery remained. The detection of the X-ray source 1E0630+178 with unusual properties in the Geminga COS-B error box with the Einstein Observatory⁹ provided the first real hope of a counterpart at another wavelength. Bignami *et al.*⁹ suggested on the basis of their observation that the source is a neutron star located at ~ 100 pc. Bignami *et al.*¹⁰ then reported the detection of an optical object inside the Einstein error circle. Halpern and Tytler¹¹ confirmed that this was a very blue object. They argued that this result and other factors support the proposal that the object is a neutron star and further that the large γ -ray to X-ray flux ratio requires that the γ -rays be produced far from the surface of the neutron star, as is required in most pulsar models.

The discovery of a 237.0974-ms period in 1E0630+178 (ref. 1) has provided compelling evidence that the source is a neutron star. It also yielded the information needed to search for periodicity in the γ -ray data; otherwise, the search for a period of this magnitude in high-energy γ -ray data is difficult or impossible because the paucity of photons leads to a substantial probability of finding chance apparent structure in the phase diagram. On being informed of the periodicity and upper limit to the period derivative (J. P. Halpern and S. S. Holt, personal communication), we made a search of the high-energy γ -ray data already obtained by the EGRET.

The EGRET on the Compton Gamma-Ray Observatory has the typical components used in high-energy γ -ray telescopes: an anticoincidence system to discriminate against charged particle radiation, a spark chamber system with interspersed pair conversion material to convert the γ -ray and determine the trajectories of the secondary electron-positron pair, a triggering telescope which detects the presence of charged particles with the correct direction of motion, and an energy-measuring device, which in the case of EGRET is a NaI(Tl) crystal. Descriptions and general capabilities of the instrument are given in refs 12–14. The telescope covers an energy range from ~ 20 MeV to >20 GeV with more than an order of magnitude greater sensitivity than the SAS-2 or COS-B instruments, as well as improved resolution

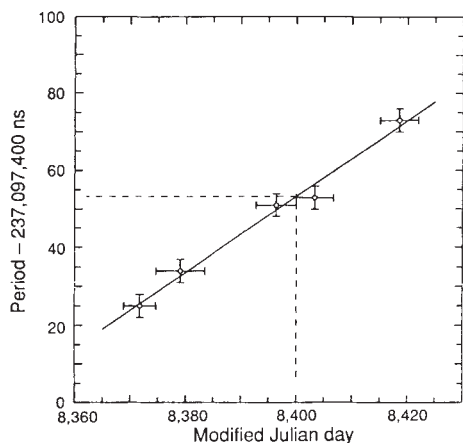


FIG. 1 Observed period as a function of the modified Julian day. The line is a fit of the form $P \times (\text{day} - 8,400) + P_0$, where P_0 is the period at epoch 8,400. The values of P and P_0 are quoted in the text.

in energy, angle and timing measurements. Because of the very low flux level of the high-energy γ -rays, observing periods are typically ~ 2 weeks.

During the Compton Observatory observations on 22 April–7 May, 16–30 May and 8–15 June 1991, EGRET viewed a region encompassing the galactic anticentre. A principal target was Geminga, and preliminary results showed the suggested X-ray counterpart 1E0630+178 to be the closest identified object to the EGRET source location¹⁵. We selected a sample of data in the energy range 100–500 MeV from the 16–30 May EGRET observation. On the basis of the Rosat observation ~ 63 days earlier, a scan with a zero value of the period derivative over the range 0.237097–0.237099 s produced a strong signal ($\chi^2 = 400$ for 19 degrees of freedom) near the period quoted below. The probability of chance coincidence, considering the number of trials, is virtually zero. We then searched the other two observation periods within a much narrower range of periods to confirm the pulsed nature of the emission.

The period derivative was measured by dividing the data into ~ 5 -day intervals and searching each in a restricted range. This process was repeated after the first determination of the derivative to obtain the results given in Fig. 1. In each search, the period that maximized the χ^2 statistic was used. To make the Earth–Sun barycentre correction, we used the J2000 coordinates of 1E0630+178 ($\alpha = 06$ h 33 min 53.9 s and $\delta = 17^\circ 46' 10.6''$).

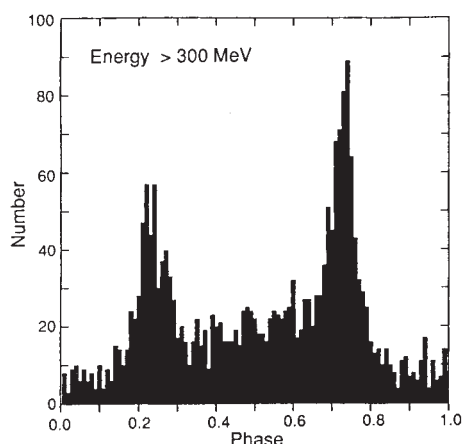


FIG. 2 Light curve of γ -rays ($E > 300$ MeV) for the combined observation periods 22 April–7 May, 16–30 May and 8–15 June 1991. The epoch, period and period derivative values quoted in the text were used.

The EGRET period extrapolated back to the epoch of the Rosat observation is in excellent agreement with the period found by Halpern and Holt¹.

The timing information, derived from the EGRET data, is $t_0 = \text{JD (Julian day)} 2,448,400$, $P = 0.237097453 \pm 0.000000003$ s, $\dot{P} = (11.4 \pm 1.7) \times 10^{-15} \text{ s s}^{-1}$.

Figure 2 shows the arrival times of 2,211 γ -rays with energy above 300 MeV and within 3° of the known 1E0630+178 direction, folded using the derived timing parameters. In contrast to the single broad peak seen in the X-rays, two narrow peaks are evident in the γ -ray phase plot, with separation of 0.50 ± 0.03 . The 0.5 phase separation might suggest an alias, but there is strong evidence for excess γ -radiation in one of the two regions between the peaks, ruling out a period value one-half as large. Because of an uncertainty, at present unresolved, in the absolute time of the Compton Observatory data stream, the phase relationship between the X-ray pulse and the γ -ray pulses cannot yet be determined. 1E0630+718 is similar to the Crab and Vela pulsars, both of which show a double pulse, although in those pulsars the phase separation is ~ 0.4 .

The derived timing parameters are characteristic of an isolated radio pulsar, although surveys have detected no radio pulsar in this direction¹⁶. From the standard equations for the properties of a rotating, magnetized neutron star (see ref. 16), the period and period derivative give the following parameters: characteristic age $\tau = 3.2 \times 10^5$ yr; surface magnetic field $B = 1.6 \times 10^{12}$ gauss; energy loss rate $\dot{E} = 3.5 \times 10^{34} \text{ erg s}^{-1}$.

Assuming that the observed γ -radiation is powered by the spin-down energy loss of the neutron star, we can use the value of $\dot{E}$ to set a limit on the distance to the source. The observed flux from Geminga above 100 MeV is $\sim 4 \times 10^{-6} \text{ photons cm}^{-2} \text{ s}^{-1}$. The spectrum is relatively flat, with a mean energy of 300 MeV. If all the spin-down energy appears in the form of γ -rays and the beaming factor is taken to be unity, then an absolute upper limit to the distance to the source is ~ 380 pc. A fraction of 1% of the rotational energy loss appearing as γ -rays (as is seen from the Vela pulsar) would put the distance at ~ 38 pc. Geminga may, therefore, be one of the closest known neutron stars.

Gamma-ray emission from fast pulsars can be interpreted as radiation from particles accelerated in the pulsar magnetosphere, either near the polar cap surface¹⁷ or in vacuum gaps in the outer magnetosphere¹⁸. In the polar cap model, the two peaks could be interpreted as coming from the opposite poles of a neutron star whose rotational and magnetic dipole axes are nearly orthogonal. The absence of a radio pulse might be explained as either a very narrow radio beam or an absorption effect near the star. The outer gap model predicts that Vela-type pulsars with periods longer than ~ 130 ms should have their vacuum gaps quenched and hence be incapable of γ -ray production¹⁹. The present observation seems to be inconsistent with this model. $\square$

Received 22 April; accepted 7 May 1992.

- Halpern, J. P. & Holt, S. S. *Nature* **357**, 222–224 (1992).
- Kniffen, D. A. et al. *Proc. 14th Int. Cosmic Ray Conf.* Vol. 1, 100–105 (MPE, Munich, 1975).
- Fichtel, C. E. et al. *Astrophys. J.* **198**, 163–182 (1975).
- Thompson, D. J. et al. *Astrophys. J.* **213**, 252–262 (1977).
- Bennett, K. et al. *Astr. Astrophys.* **56**, 469–471 (1977).
- Masnou, J. L. et al. *Proc. 17th Int. Cosmic Ray Conf.* Vol. 1, 177–180 (CEN Saclay, Paris, 1981).
- Hermesen, W. *EGRET Science Symp.* NASA CP 3071, 161–177 (NASA/GSFC, Greenbelt, Maryland, 1990).
- Spolestra, T. A. Th. & Hermesen, W. *Astr. Astrophys.* **135**, 135–140 (1984).
- Bignami, G. F., Caraveo, P. A. & Lamb, R. C. *Astrophys. J.* **272**, L9–L13 (1983).
- Bignami, G. F., Caraveo, P. A., Paul, J. A., Salotti, L. & Vigroux, L. *Astrophys. J.* **319**, 358–361 (1987).
- Halpern, J. P. & Tytler, D. *Astrophys. J.* **330**, 201–217 (1988).
- Hughes, E. G. et al. *IEEE Trans. Nucl. Sci.* NS-27, 364–369 (1980).
- Kanbach, G. et al. *Space Sci. Rev.* **49**, 69–84 (1988).
- Kanbach, G. *Gamma Ray Observatory Science Workshop Proc.* 2-1 to 2-10 (NASA/GSFC, Greenbelt, Maryland, 1989).
- Mayer-Hasselwander, H. A. et al. *Astr. Astrophys.* (in the press).
- Manchester, R. N. & Taylor, J. H. *Astr. J.* **86**, 1953–1973 (1981).
- Daugherty, J. K. & Harding, A. *Astrophys. J.* **252**, 337–347 (1982).
- Cheng, K. S., Ho, C. & Ruderman, M. *Astrophys. J.* **300**, 500–521 (1986).
- Ruderman, M. & Cheng, K. S. *Astrophys. J.* **335**, 306–318 (1988).

Origin of the hot gas and radio blobs at the Galactic Centre

Mark Wardle & Farhad Yusef-Zadeh

Department of Physics and Astronomy, Northwestern University,
2145 North Sheridan Road, Evanston, Illinois 60208, USA

RECENT infrared observations¹ have revealed hot gas, with a temperature possibly as high as one million degrees, associated with a small cavity 2 arcseconds in diameter in one of the ionized gas streamers (the 'Bar') that orbit the Galactic Centre radio source SgrA* (ref. 2), thought to contain a massive black hole. Radio continuum observations³ show a chain of blobs of emission leading from SgrA* to the small cavity. We present here further high-resolution radio images which show that the blobs are connected to SgrA* by a ridge of emission. We suggest that the blobs are formed by the interaction of stellar winds from the IRS16 cluster with the gravitational potential of SgrA*. The hot gas¹ then results from the dissipation of the kinetic energy of the blobs as they collide with the orbiting ionized streamer. These collisions are of dynamical significance for the motion of the Bar around the Galactic Centre, and there should be detectable variability in the structure on a timescale of 10 years.

Figure 1 shows a schematic diagram of the inner light year of the Galaxy. SgrA* is a bright, compact radio source located very close to the dynamical centre of the Galaxy. It is variable, shows nonthermal characteristics with a spectrum resembling the cores of extragalactic radio sources² and is a massive black-hole candidate. A mass of 3×10^6 solar masses, $M_\odot$, is suggested by the velocities of the ionized gas orbiting the Galactic Centre^{4,5} and of the late-type stars in the cluster that engulfs SgrA* (refs 6, 7). A cluster of at least 20 infrared sources, IRS16, lies at 3" to the east of SgrA* (ref. 1). The line emission arising from the IRS16 subcomponents has a spectrum which is consistent with that of mass-losing WN stars⁸. Indeed, IRS16 is the source of a powerful wind. Broad He I, Br α , and Br γ emission lines from the vicinity of IRS16 indicate an outflow with a terminal velocity of 500–1,000 km s⁻¹ (refs 8–13), with estimates of the total mass-loss rate ranging between $10^{-3.5}$ and $10^{-2.5} M_\odot$ yr⁻¹. The ram pressure of the wind is probably responsible for the ionized cometary tail of the supergiant IRS7 (refs 14–16). Also, there is morphological evidence for the interaction of the wind with the northern arm of SgrA West and the circumnuclear ring¹⁷.

The detection of 2.217- μ m line emission from hot gas filling the small cavity¹⁸ in the Bar has recently been reported¹. The cavity and the location of the 2.217- μ m emission are sketched in Fig. 1. This result bears on the formation of the cavity and provides a clue as to the sources and nature of the activity at the Galactic Centre. A supernova explosion occurring within the ionized stream can be ruled out as a cause¹ because the nonthermal emission characteristic of young supernova remnants is missing and because the hot gas is not moving fast enough relative to the ambient material. Direct interaction with the stellar winds from two prominent infrared sources, the IRS16 cluster and IRS13, can be excluded because the cavity is closed off in those directions. In ref. 1 it is suggested that a wind from SgrA* is a likely energy source, although there are difficulties with the most straightforward application of this idea. SgrA* is located at the northeastern periphery of the small cavity and it is difficult to see how the cavity is produced from a displaced source of wind¹⁸. As pointed out in ref. 1, the wind must be anisotropic because the cavity is fairly compact and appears partially closed off towards SgrA*.

An important clue to the origin of the small cavity is the chain of several blobs of thermal radio continuum emission that traces out an arc linking SgrA* and the cavity in the Bar³. Zhao *et al.*¹⁹ have recently noted that the cavity is slightly elongated towards the southeast. The blobs (sources ϵ , ζ , η in ref. 3) have

been detected at a number of frequencies¹⁹ and their reality is well established. Figure 2a and b shows further images of the radio blobs at 3.5 and 6 cm wavelengths, with resolutions of $0.5'' \times 0.2''$ and $0.6'' \times 0.35''$, respectively. These radio images are constructed from observations made with the Very Large Array. A detailed account of the 6-cm observations is given in ref. 3; the 3.5-cm observations are part of an ongoing mosaic imaging of the SgrA complex and will be described in more detail elsewhere. We note that the blobs ϵ , ζ and η break up into a number of components. The striking morphological link to SgrA* is best illustrated in Fig. 2a as a ridge of emission associated with the blobs. This suggests that the small cavity and its southeastern extension are a result of one or more blobs running into the Bar, producing the hot shocked gas responsible for the 2.217- μ m emission. Here we suggest that the 'SgrA* wind' is a portion of the wind from the stellar cluster IRS16 that has been collimated by the gravitational potential of a massive black hole coincident with SgrA*, and that radio blobs may be formed as a result of this interaction. This is an extension of the idea that the luminosity of SgrA* is provided by the accretion of part of the wind from the IRS16 cluster^{20,21}.

What do the 2.217- μ m observations imply? Unfortunately, the line is not yet identified; it is likely to be either the $^2D_{5/2} \rightarrow ^2D_{3/2}$ transition of Fe XII or the $^3G_5 \rightarrow ^3H_6$ transition of Fe III. In the former case the hot gas has a temperature T_h in the range $(1-3) \times 10^6$ K, and a total mass of hot gas $M_h \approx 0.1$ (4×10^4 cm⁻³/ n_e) $M_\odot$, where $M_\odot$ is the solar mass and n_e is the electron number density¹. Gas in this temperature range cools at a rate $\sim 1 \times 10^{-22} n_e n_H$ erg cm⁻³ s⁻¹, where n_H is the hydrogen number density in units of cm⁻³ (ref. 22). The cooling time for $n_e = 4 \times 10^4$ cm⁻³ and $T_h = 2 \times 10^6$ K is 6 years; energy is lost by the hot gas at a rate $L_h \approx 4 \times 10^{38}$ erg s⁻¹. Observations of the Galactic Centre with the Imaging Proportional Counter of the Einstein Observatory with a spatial resolution of ~ 1 arcmin have shown an X-ray source 1E1742.5–2859 with a luminosity between 0.5 and 4.5 keV of 2×10^{35} erg s⁻¹ (assuming a distance of 8.5 kpc and $A_V = 30$) coincident with SgrA West²³. This source

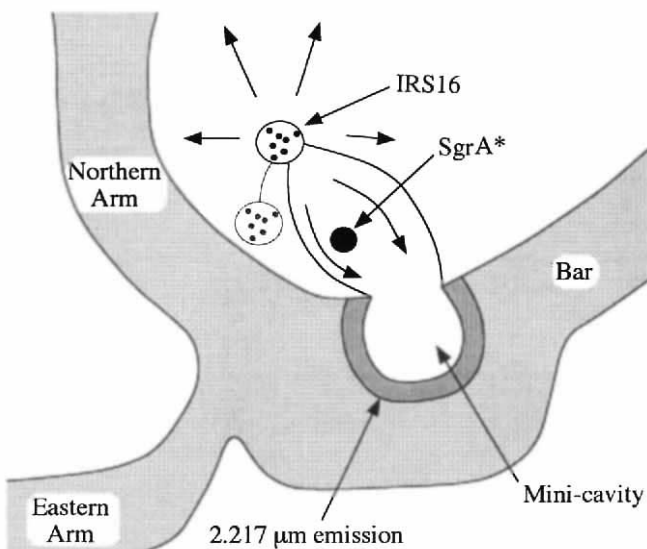


FIG. 1 Schematic diagram of a model of the inner light year of the Galactic Centre region. SgrA* is a massive black hole which provides the gravitational anchor for the region. IRS16 is a cluster of mass-losing stars whose winds combine to form an overall outflow. A portion of the outflow is gravitationally collimated by SgrA* and collides with one of the ionized streamers in orbit around SgrA* to form a cavity. The collision drives shocks into the streamer; shock heated gas is responsible for the 2.217- μ m emission. The orbital motion of IRS16 that may produce some of the curvature in the collimated outflow is also indicated.

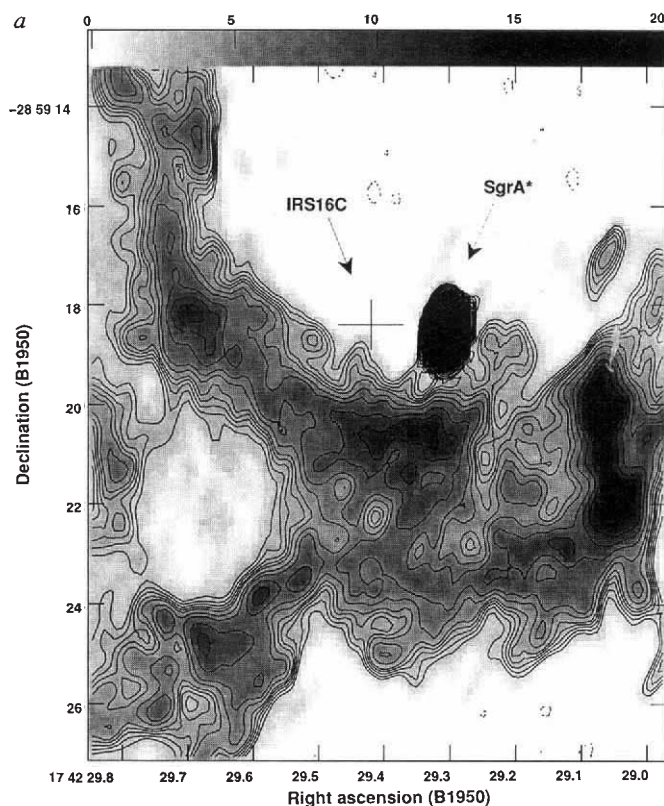
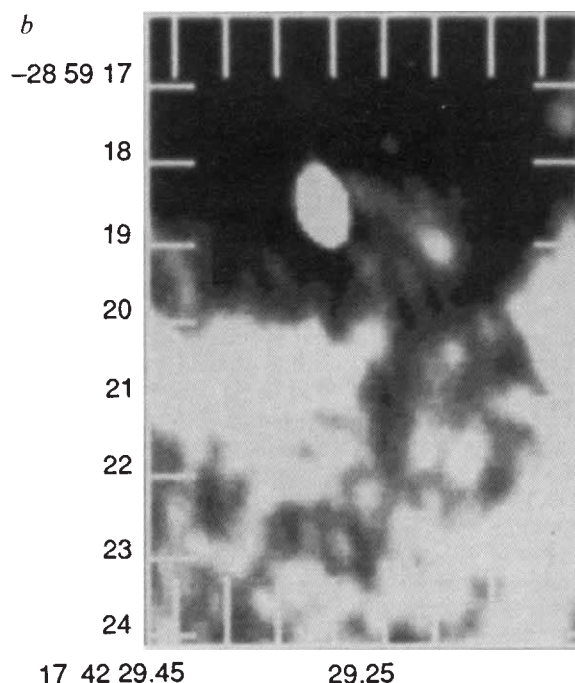


FIG. 2 *a*, A radiograph at 3.5 cm wavelength of the region around SgrA*, with a spatial resolution of $0.5'' \times 0.2''$. The cavity in the ionized gas and the blobs connecting it to SgrA* can also be seen. *b*, Contours of total intensity at 6 cm wavelength with a resolution of $0.62'' \times 0.35''$ are set at



–2, 2, 2.5, 3, 3.5, 4, 5, 6, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35 and 40 mJy per beam area. The cross represents the position of IRS16C including its positional uncertainties²⁸.

may be identified with the hot gas associated with the cavity¹. The X-ray luminosity in this energy range, from $0.1 M_{\odot}$ of gas at 2.5×10^6 K with $n_e = 4 \times 10^4 \text{ cm}^{-3}$ and solar abundances, is $\sim 2 \times 10^{37} \text{ erg s}^{-1}$ (ref. 22), two orders of magnitude larger than that inferred by the Einstein observations. This discrepancy could be partially resolved if (as is likely) the metallicity of the gas is higher than solar, so that the mass implied by the 2.217- μm observations is reduced, but it should be noted that most of the X-ray flux is in O VII lines²² so that the ratio of 2.217- μm flux to X-ray flux scales as $[\text{Fe}/\text{O}]$ rather than $[\text{Fe}/\text{H}]$. A more likely explanation for the discrepancy is that the local extinction towards the hot bubble is $A_V = 50$, which would increase the luminosity inferred from the Einstein observations by two orders of magnitude. If Fe III is responsible for the 2.217- μm emission then the mass of gas is similar with a temperature between 40,000 and 10^5 K (ref. 1). The cooling rate is between 1 and 6 times larger, and the X-ray emission is negligible.

The brightness temperature of the blobs is in the range 500–1,000 K, whereas the electron temperature should be close to 10^4 K. The radio continuum emission is therefore optically thin and the emission measure and blob size (diameter around $1''$) then imply a mass $M_b \approx 2 \times 10^{-2} M_{\odot}$, a blob column density of $3 \times 10^{21} \text{ cm}^{-2}$ and a number density of $\sim 5 \times 10^4 \text{ cm}^{-3}$. The curvature of the chain could either reflect the trajectories of the blobs or be the result of the orbital motion of IRS16 (if the blobs are a result of the focusing of the wind by SgrA*). In either case, the blob velocity is roughly the escape velocity from SgrA*, which is 500 km s^{-1} at 0.1 pc. The timescale for a blob to traverse the length of the chain is ~ 100 yr for a blob velocity of $1,000 \text{ km s}^{-1}$, so the mass-loss rate of SgrA* in the form of blobs is roughly one blob every 30 years or $6 \times 10^{-4} M_{\odot} \text{ yr}^{-1}$. Adopting a relative velocity between blob and bar of $v_b = 1,000 \text{ km s}^{-1}$, the kinetic energy of the blob is $0.5 M_b v_b^2 \approx$

$2 \times 10^{47} \text{ erg}$. The column density in the bar is similar to that in the blobs, so the blobs are slowed on a timescale $(M_b v_b) / (\pi \rho_0 r_b^2 v_b^2)$, where ρ_0 is the density of the Bar. This is roughly 25 yr for a blob radius $r_b = 0.03 \text{ pc}$ and a bar number density of 10^{-4} cm^{-3} , yielding an energy loss rate of $3 \times 10^{38} \text{ erg s}^{-1}$, which is comparable to the thermal losses of the hot gas. The stopping timescale is comparable to the time between blob collisions, so that the hot gas can be replenished in a quasi-steady state. The postshock gas temperature in a plane-parallel shock is $1.4 \times 10^7 (v_s/1,000 \text{ km s}^{-1})^2 \text{ K}$ where v_s is the shock velocity; it is therefore plausible that the Bar material is heated to a few million degrees on passing obliquely through the bow shock caused by the passage of the blob. The energetics therefore seem consistent if Fe XII is producing the 2.217- μm emission. The Fe III case is problematic because it is unlikely that the postshock temperature is as low as 10^5 K. Then gas must be shock-heated to 10^6 K at a rate which is a factor $t_c(10^6 \text{ K}) / t_c(10^5 \text{ K})$ faster than in the Fe XII case, where t_c is the cooling time. Even neglecting the increasing density as the gas cools, this ratio is 60, although heating by the ultraviolet radiation field at the Galactic Centre may partly offset this.

How are the blobs formed and ejected from SgrA*? Because the chain lies opposite IRS16, we suggest that the blobs are formed as a result of the gravitational collimation of the combined stellar winds of IRS16 by a mass of $3 \times 10^6 M_{\odot}$ located at the position of SgrA*. If we adopt a mass loss in the wind of $\dot{M}_w = 3 \times 10^{-3} M_{\odot} \text{ yr}^{-1}$, one-fifth of this must be collected into blobs. The IRS16 wind velocity, $v_w \approx 700 \text{ km s}^{-1}$, is comparable to the escape velocity from the potential, suggesting that the collimation could be considerable. One possibility is that the blobs are formed as the byproduct of a dynamical instability in the accretion shock²⁴. But the accretion rate onto the black hole is then likely to be at least as large as the mass-loss rate in the

form of blobs, $\dot{M}_b$, and the accretion luminosity is $\epsilon \dot{M}_b c^2 \approx \epsilon \times 10^{10} L_\odot$ where $L_\odot$ is the solar luminosity. This is too large unless the efficiency is rather low, $\sim 0.1\%$ or less. An alternative is that the blobs are condensations resulting from a thermal instability in the wind triggered by the gravitational focusing of SgrA*. In this picture the individual stellar winds from cluster members are shock-heated to 10^6 K as they merge to form the overall outflow. Adopting a spherically symmetric constant-velocity model for the wind, we find that the density at 0.1 pc from the centre (corresponding to the displacement of SgrA*) is $\sim 10^3 \text{ cm}^{-3}$. The scale of condensations is set by the requirement that the sound crossing time be comparable to the cooling time²⁵, yielding a size $\sim 5 \times 10^{16}$ cm.

The collisions should have dynamical consequences, and indeed, the ionized gas associated with the Bar is known to be kinematically disturbed when compared with ionized gas associated with the streamers^{4,26,27}. In particular, a pair of high-velocity clouds (at radial velocities of +260 and -300 km s⁻¹) which are coincident with the western periphery of the cavity and its southeastern extension do not seem to fit a single orbit associated with the ionized streamers⁴, and the velocity gradient across them is larger by a factor of two. One would expect the interaction of the remaining part of the wind with the ionized streamers to be less violent because the wind material is not collected into dense blobs. In fact, close to IRS16, there is a discontinuity in the velocity gradient along the northern arm (near IRS1; ref. 4) where the western edge becomes sharp and radio-bright¹⁸. The disturbed kinematics of the Bar and northern arm may affect estimates of the mass of the black-hole candidate that are based on the velocity distribution of ionized gas in the inner few arcseconds of the Galaxy.

Our high-resolution radio images of the Galactic Centre provide further morphological evidence that the blobs are associated with SgrA*. The model that we have proposed—that the blobs are produced by interaction of stellar winds with the gravitational potential of a massive black hole—predicts variability on a timescale of 10 yr and can be tested by proper-motion studies of the blobs at radio wavelengths. $\square$

Received 2 January; accepted 3 April 1992.

1. Eckart, A. et al. *Nature* **355**, 526–529 (1992).
2. Lo, K. Y. *IAU Symp. 136, The Center of the Galaxy* (ed. Morris, M.) 527–534 (Kluwer, Dordrecht, 1989).
3. Yusef-Zadeh, F., Morris, M. & Ekers, R. *Nature* **348**, 45–47 (1990).
4. Serabyn, E., Lacy, J. H., Townes, C. H. & Bharat, R. *Astrophys. J.* **326**, 171–185 (1988).
5. Lacy, J. H., Achtermann, J. M. & Serabyn, E. *Astrophys. J.* **380**, L71–L74 (1991).
6. McGinn, M. T., Sellgren, K., Becklin, E. E. & Hall, D. N. B. *Astrophys. J.* **338**, 824–840 (1989).
7. Sellgren, K., McGinn, M. T., Becklin, E. E. & Hall, D. N. B. *Astrophys. J.* **359**, 112–120 (1990).
8. Allen, D. A., Hyland, A. R. & Hillier, D. J. *Mon. Not. R. astr. Soc.* **244**, 706–713 (1990).
9. Hall, D. N. B., Kleinmann, S. G. & Scoville, N. Z. *Astrophys. J.* **260**, L53–L57 (1982).
10. Geballe, T. R., Krisciunas, K., Lee, T. J., Gatley, I. & Wade, R. *Astrophys. J.* **284**, 118–125 (1984).
11. Maillard, J. P. & Gay, J. *IAU Symp. 136, The Center of the Galaxy* (ed. Morris, M.) 517–521 (Kluwer, Dordrecht, 1989).
12. Geballe, T. R., Krisciunas, K., Bailey, J. A. & Wade, R. *Astrophys. J.* **370**, L73–L76 (1991).
13. Brabbe, A., Genzel, R., Drapatz, S. & Rotaciuc, V. *Astrophys. J.* **382**, L19–L22 (1991).
14. Yusef-Zadeh, F. & Morris, M. *Astrophys. J.* **371**, L59–L62 (1991).
15. Serabyn, E., Lacy, J. H. & Achtermann, J. M. *Astrophys. J.* **378**, 557–564 (1991).
16. Yusef-Zadeh, F. & Melia, F. *Astrophys. J.* **385**, L41–L44 (1992).
17. Yusef-Zadeh, F. & Wardle, M. In *The Large Scale Distribution of Stars and Gas in the Milky Way* (ed. Blitz, L.) (Kluwer, in the press).
18. Yusef-Zadeh, F., Morris, M. & Ekers, R. *IAU Symp. 136, The Center of the Galaxy* (ed. Morris, M.) 443–451 (Kluwer, Dordrecht, 1989).
19. Zhao, J.-H., Goss, W. M., Lo, K. Y. & Ekers, R. D. *Nature* **354**, 46–48 (1991).
20. Ozernoy, L. M. *IAU Symp. 136, The Center of the Galaxy* (ed. Morris, M.) 555–566 (Kluwer, Dordrecht, 1989).
21. Melia, F. *Astrophys. J.* **387**, L25–L28 (1992).
22. Gaetz, T. J. & Salpeter, E. E. *Astrophys. J. (Suppl.)* **52**, 155–168 (1983).
23. Watson, M. G., Willingale, R., Grindlay, J. E. & Hertz, P. *Astrophys. J.* **250**, 142–154 (1981).
24. Taam, R. E. & Fryxell, B. A. *Astrophys. J.* **327**, L73–L75 (1988).
25. Field, G. B. *Astrophys. J.* **142**, 531–567 (1965).
26. Brown, R. L. & Liszt, H. H. A. *Rev. Astr. Astrophys.* **22**, 223–265 (1984).
27. Roberts, D. thesis, Univ. of Oklahoma (1992).
28. Tollestrup, E. V., Capps, R. W. & Becklin, E. E. *Astr. J.* **98**, 204–216 (1989).

ACKNOWLEDGEMENTS. The NRAO is operated by Associated Universities, Inc. under a cooperative agreement with the NSF. This work was supported in part by NASA. M. Wardle is supported by the Lindheimer Fellowship at Northwestern University. We thank R. Genzel, M. Morris, and R. Taam for comments and discussions.

Multiple-shell structures of laser-cooled $^{24}\text{Mg}^+$ ions in a quadrupole storage ring

G. Birkel, S. Kassner & H. Walther

Max-Planck-Institut für Quantenoptik, Garching bei München, Germany

THE possibility of creating ordered ion beams in high-energy storage rings^{1,2} by means of electron and laser cooling has opened up a new era in accelerator physics. The enhanced luminosity and suppressed momentum spread in such systems create the highest possible phase-space density. The first experimental results were obtained by cooling $^7\text{Li}^+$ beams to temperatures of a few kelvin or even to sub-kelvin temperatures^{3,4}, and the ordered structures have been studied theoretically^{5–7} by methods of molecular dynamics. Predicted configurations for the lowest ion densities have been observed in low-energy quadrupole storage rings⁸ and linear traps⁹. Recently we showed that at slightly higher ion densities helical structures are obtained¹⁰. Here we present a series of new experimental results on ordered ion structures in a quadrupole storage ring. In order of increasing ion number, a linear chain of ions, a zig-zag structure, helical structures and finally multiple concentric shells could be observed. The experimental results agree with molecular dynamics calculations.

When ordered structures in storage rings are simulated^{5–7}, a cylindrically symmetric, static harmonic potential is usually assumed to describe the confining field. The pseudopotential of a low-energy quadrupole storage ring is closer to the theoretical model than the confining field of a high-energy storage ring, where the ions are subjected to periodic squeezing and pulling in the focusing sections, Coulomb explosion in the drift sections and shear forces in the bending sections. The principle and first realizations of quadrupole storage rings are described in refs 11–13. In the ring the ions can move freely along the axis of the quadrupole field and are radially confined by applying a radiofrequency (r.f.) voltage $U_{\text{RF}} \cos \Omega t$ to the electrodes. This leads to a harmonic confining potential¹⁴ of the form $\psi = \psi_0 r^2 / r_0^2$ with the potential depth $\psi_0 = q^2 U_{\text{RF}}^2 / 4m\Omega^2 r_0^2$, where q is the ion charge, m the ion mass, r the distance from the field axis and r_0 half the distance between opposite electrodes. The oscillation frequency in this harmonic potential, $\omega_{\text{sec}} = (2\psi_0 / mr_0^2)^{1/2}$,

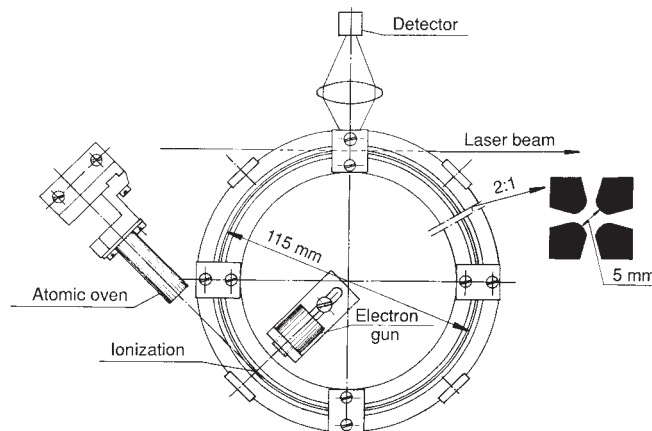


FIG. 1 Quadrupole storage ring, with the atomic beam oven and electron gun. The storage ring consists of four circular electrodes, and the diameter of the toroidal storage volume is $2R = 115$ mm. The insert shows an enlarged cross-section with opposite electrodes having a separation of $2r_0 = 5$ mm. The laser beam enters the storage volume tangentially. Resonance fluorescence is detected with a photomultiplier tube or an imaging photon detector system.

FIG. 2 Colour-coded images of crystal-line structures of laser-cooled $^{24}\text{Mg}^+$ ions. The intensity increases from violet to blue, yellow and red. Individual ions could be resolved in these images. The ions arrange themselves in minimum energy configurations. *a*, For low ion density ($\lambda=0.29$) the ions form a string along the field axis; *b*, increasing the ion density changes the configuration to a zig-zag ($\lambda=0.92$). At still higher ion densities the ions form ordered helical structures on the surface of a cylinder: *c*, two interwoven helices at $\lambda=1.9$; *d*, three interwoven helices at $\lambda=2.6$. Experimental images are displayed above, visualizations below.

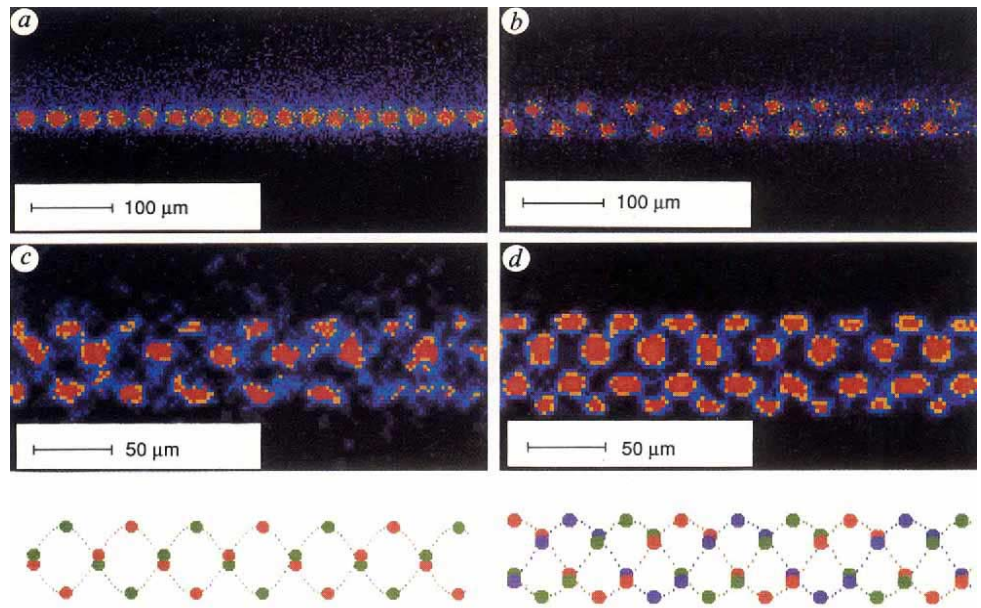
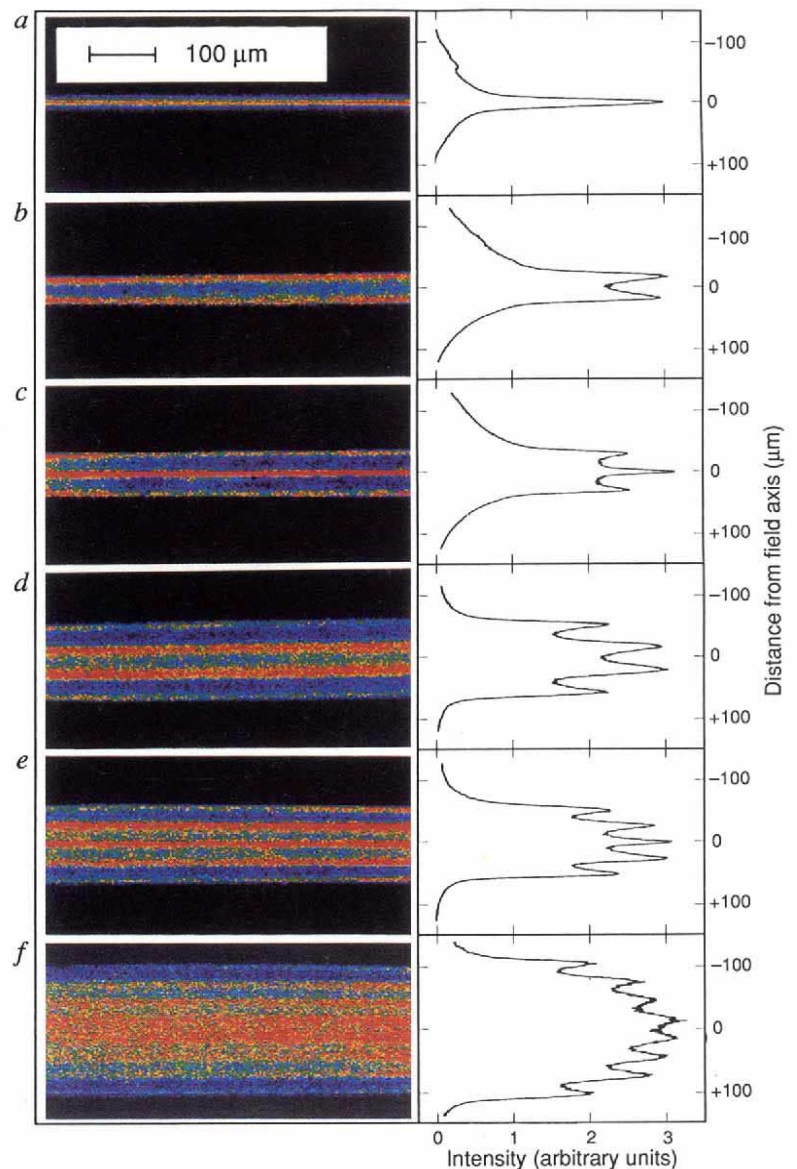


FIG. 3 Images and intensity profiles of (a) string, (b) one shell ($\rho/a=1.05$, total ion number in the ring $N \approx 5 \times 10^4$), (c) one shell plus string ($\rho/a=1.8$, $N \approx 1 \times 10^5$), (d) two shells ($\rho/a=2.7$, $N \approx 2 \times 10^5$), (e) two shells plus string ($\rho/a=3.4$, $N \approx 3 \times 10^5$) and (f) four shells ($\rho/a=6.2$, $N \approx 8 \times 10^5$). The ions are not individually resolved. The structures can be identified with the help of the radial intensity profiles (right). The images are colour-coded as in Fig. 2. Integration times are longer than 3 min each.



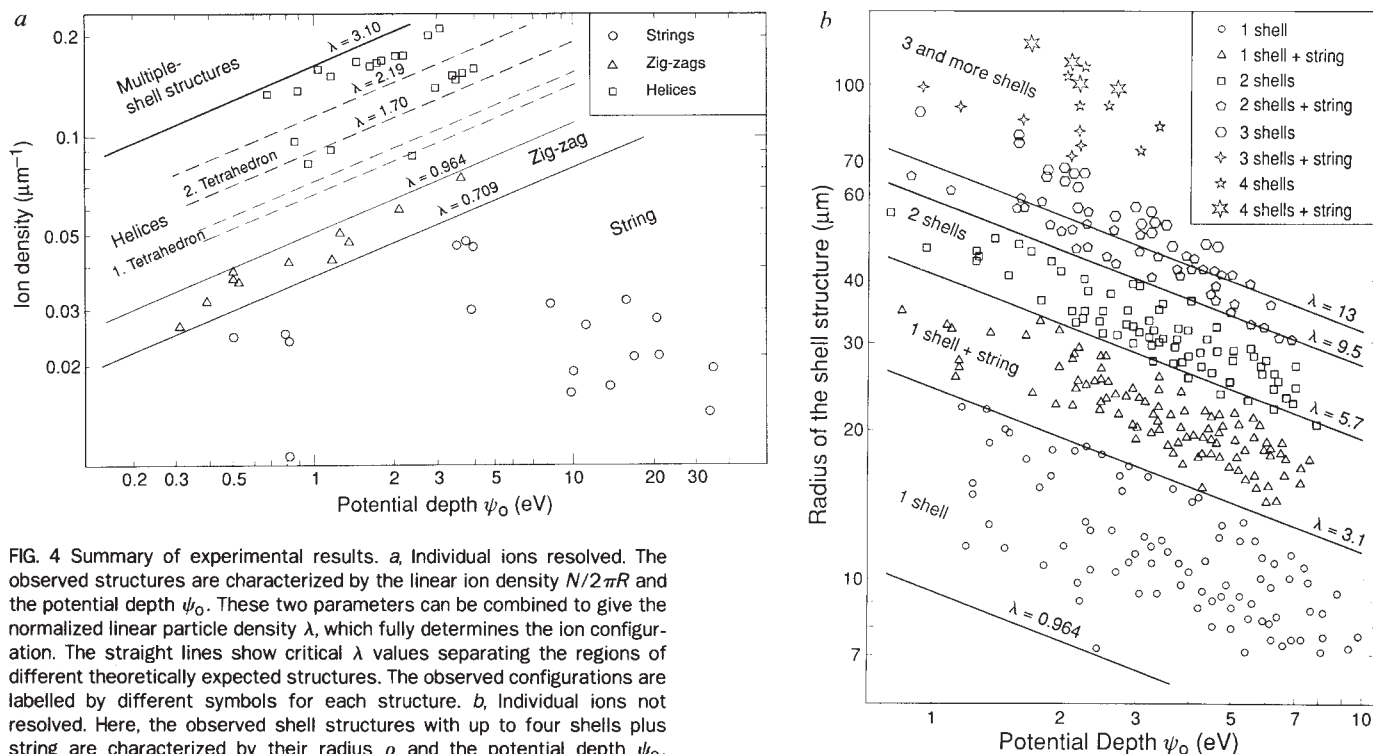


FIG. 4 Summary of experimental results. *a*, Individual ions resolved. The observed structures are characterized by the linear ion density $N/2\pi R$ and the potential depth ψ_0 . These two parameters can be combined to give the normalized linear particle density λ , which fully determines the ion configuration. The straight lines show critical λ values separating the regions of different theoretically expected structures. The observed configurations are labelled by different symbols for each structure. *b*, Individual ions not resolved. Here, the observed shell structures with up to four shells plus string are characterized by their radius ρ and the potential depth ψ_0 . Different observed structures are again separated by lines of theoretically determined critical λ .

is called the secular frequency¹⁴. Ions in a r.f. quadrupole field also show micromotion with frequency Ω , the amplitude of which increases in proportion to the distance r from the field axis. To compare experimental results with theory, it is useful to define the normalized 'linear particle density'⁷, $\lambda = aN/2\pi R$. Here N is the total number of stored ions in the ring, R the ring radius and a the Wigner-Seitz radius, $a = (3q/8\pi\epsilon_0 K)^{1/3}$, where K is the force constant. For the ring trap, $K = 2\psi_0/qr_0^2$, which leads to $a = (3q^2r_0^2/16\pi\epsilon_0\psi_0)^{1/3}$. The quantity a^3 corresponds to the volume of the unit cell of the Coulomb crystal. The equilibrium structures resulting from the molecular dynamics (MD) calculations as a function of increasing λ are a linear chain, a zig-zag, a single shell and multiple shells⁵⁻⁷. For the single shell, different ion configurations on the shell surface, such as single and multiple helices, have also been calculated in detail^{6,7}. Static properties of these configurations, such as shell diameters and the arrangement of the ions on the shell surfaces, can also be obtained by minimizing the energy in an electrostatic model^{7,15}.

Figure 1 shows the quadrupole storage ring. The toroidal trap volume has a diameter $2R = 115$ mm, and the distance between opposite electrodes is $2r_0 = 5$ mm. The radiofrequency is $\Omega = 2\pi \times 6.56$ MHz, and the secular frequency ω_{sec} is of the order of megahertz. The geometry of the quadrupole and measurement of the r.f. voltage allow the potential depth to be determined with an accuracy of 5%. The apparatus is contained in an ultra-high-vacuum chamber with a pressure below 10^{-8} Pa. Magnesium ions $^{24}\text{Mg}^+$ are created by electron bombardment of a thermal atomic beam. A contact potential barrier of ~ 1 eV is produced by deposition of stray magnesium atoms on the electrodes in the region where the atomic beam passes through the quadrupole. The potential barrier enables us to cool the stored ions to rest, using only a single laser beam¹⁰. Laser light with a wavelength of 280 nm, resonant with the $3^2\text{S}_{1/2} - 3^2\text{P}_{3/2}$ transition of $^{24}\text{Mg}^+$, is focused tangentially along the axis of the quadrupole field, where it can interact with the stored ions. The resonance line of $^{24}\text{Mg}^+$ can easily be saturated with the available laser

power of several milliwatts. The fluorescence light emitted by the ions is measured by a photomultiplier tube and an imaging photon detector system equipped with a two-stage microchannel plate and a position analyser, the channel diameter being 25 μm . The lens system used for the optical imaging has a focal length of 35 mm and allows the image of the ion structure to be magnified between 10 and 70 times. For most of the measurements a magnification factor of 40 was used. Thus the imaging system allows a resolution of ~ 1 μm in the storage volume. The diffraction-limited resolution of the optics is 0.4 μm .

Previous work has shown^{8,10} that laser Doppler cooling of the stored Mg^+ ions leads to a phase transition from a disordered state to an ordered crystalline structure with the same features as described for ions stored in a Paul trap^{16,17}. In contrast to high-energy storage rings, in the quadrupole ring the laser-cooled ions are at rest, and therefore the structures can be investigated in detail. Ordered ion configurations have also been observed in Paul¹⁶⁻¹⁸ and Penning¹⁹ traps.

Figure 2 shows ordered structures for low ion densities, leading to images of singly resolved ions. The ion spacings are known to an accuracy of 3%, and therefore the ion density can be directly determined from the images with the same error. The λ parameters can be calculated with 5% accuracy for these structures. A linear chain of ions with an ion spacing of 33 μm is presented in Fig. 2a. Increasing the ion density favours configurations in which the ions are displaced from the field axis, leading to a zig-zag structure (Fig. 2b). The ions are arranged in a plane perpendicular to the direction of observation, as can be concluded from their spacing (26 μm in the direction of the field axis) and the potential depth. Further increasing the ion density changes the equilibrium configuration into helical structures. The experimental image in Fig. 2c shows an ordered structure which can be described as two interwoven helices with a diameter of 44 μm and four ions per pitch. The image in Fig. 2d, recorded with an even higher ion density, shows a structure of three interwoven helices with a diameter of 53 μm and six ions per pitch. The fact that the structures in

Fig. 2b to d do not rotate around the field axis shows that there must be small deviations from the rotational symmetry around the quadrupole axis, which may be caused by the contact potential or a misalignment of the trap electrodes. By increasing the radiation pressure the structures can be slightly compressed in the direction of the incoming laser beam. As mentioned above, the contact potential prevents the structures from starting to rotate around the ring.

Increasing the ion number and the potential depth accordingly leads to identical structures, but with smaller ion spacings, so that the ions can no longer be individually resolved (Fig. 3). Information about the structures can be extracted from the radial intensity profiles, giving a measure of the ion distribution in this direction. Instead of the parameter λ , it is now possible to use the radius ρ of the structure in units of the Wigner-Seitz radius a (where a is determined by the potential depth) to characterize the structures. The zig-zag and helix both produce a double-peak structure in the radial intensity profile (Fig. 3b). If, therefore, the ion density is so high that observation of individual ions is not possible, we cannot discriminate between the zig-zag and helical structures. This is only true because we are looking perpendicularly to the plane of the zig-zag, as in Fig. 2b. Successively increasing the ion density or lowering the potential leads to the occurrence of a string inside the shell (Fig. 3c), two concentric shells (Fig. 3d) and two shells plus string (Fig. 3e). Figure 3f shows a four-shell structure for very high densities. We have observed all possible shell structures up to four shells plus string. All these structures form after a phase transition from a disordered state, indicating that the ions occupy fixed positions within the shells. This differs from the observations of shell structures in a Penning trap¹⁹. The structures in Figs 2b–d and 3b–f are not free of micromotion, but the amplitude of this motion is too small to be observable at the present optical resolution.

The intensity profiles of Fig. 3 show decreasing contrast for a higher number of shells. This is partly due to the summation of the intensity distributions of more and more shells in a multiple-shell configuration, which masks the individual structures. In addition, for higher-order structures the inner shells become substantially thicker, as pointed out in ref. 7. Our experiments show that it is possible to obtain ordered configurations even for very high ion densities.

Figure 4a gives a summary of experimental data for all recorded images in which the ions were individually resolved. The potential depth ψ_0 and linear ion density $N/2\pi R$ (obtained directly from the images with the known optical magnification) are the experimental parameters. The theoretical limits⁷ between the different structures are given by the straight lines with constant λ . String structures are expected for $\lambda < 0.709$, zig-zag structures in the range $0.709 < \lambda < 0.964$ and single shells in the range $0.964 < \lambda < 3.10$. In the single-shell regime, many different structures are expected that are practically degenerate in energy. The structures may change on a timescale short compared with the averaging time of the imaging system, owing to their residual thermal energy. For that reason it is difficult to obtain clear images in much of this λ regime. The stability could be improved by cooling the ions further. We obtained stable configurations for several seconds (more than 10^5 secular periods of the system) near $\lambda = 1.3$ and $\lambda = 2.0$, yielding structures of two interwoven helices (Fig. 2c) and near $\lambda = 3.0$ structures of three interwoven helices (Fig. 2d). The two-fold helix is also expected from the MD calculations, the structure forming a series of tetrahedra⁷. The expected λ domains for this configuration are marked by the dashed lines in Fig. 4a. For a large range of experimental parameters the observed structures fit into the expected scheme, confirming the theoretical results. Thus the kind of structures observed is fully determined by λ , which proves to be a universal parameter independent of the experimental conditions.

Figure 4b summarizes the experimental results on the ordered shell structures up to four shells plus string without resolution

of individual ions. Again the potential depth ψ_0 is one of the experimental parameters. As the ion density can no longer be determined directly from the images, the radius ρ of the structures is used instead as the second parameter. The theoretical limits between the different shell structures are again given as straight lines with constant λ , as predicted in ref. 7, where the functional dependence of ρ/a on λ is established. The observed ion configurations are fully determined by ρ/a and therefore by λ for a variety of potential depths and ion densities. $\square$

Received 7 February; accepted 14 April 1992.

- Schiffer, J. P. & Kienle, P. Z. Phys. **A321**, 181 (1985).
- Proc. Workshop on Crystalline Ion Beams (eds Hasse, R. W., Hofmann, I. & Liesen, D.) 2–63 (GSI, Darmstadt, 1989).
- Schröder, S. et al. Phys. Rev. Lett. **64**, 2901–2904 (1990).
- Hangst, J. S. et al. Phys. Rev. Lett. **67**, 1238–1241 (1991).
- Rahman, A. & Schiffer, J. P. Phys. Rev. Lett. **57**, 1133–1136 (1986).
- Habs, D. in *Lecture Notes in Physics* 296: *Frontiers of Particle Beams* (Springer, New York, 1988).
- Hasse, R. W. & Schiffer, J. P. Ann. Phys. **203**, 419–448 (1990).
- Walther, H. in *Proc. Workshop on Light Induced Kinetic Effects on Atoms, Ions and Molecules* (eds Moi, L., Gozzini, S., Gabbanini, C., Arimondo, E. & Strumia, F.) 261–276 (ETS Editrice, Pisa, 1991).
- Raizen, M. G., Gilligan, J. M., Bergquist, J. C., Itano, W. M. & Wineland, D. J. J. mod. Opt. **39**, 233–242 (1992).
- Waki, I., Kassner, S., Birkel, G. & Walther, H. Phys. Rev. Lett. **68**, 2007–2010 (1992).
- Drees, J. & Paul, W. Z. Phys. **180**, 340–361 (1964).
- Church, D. A. J. appl. Phys. **40**, 3127–3134 (1969).
- Deutch, B. I. et al. Phys. Script. **T22**, 248–255 (1988).
- Dehmelt, H. in *Advances in Atomic and Molecular Physics* Vol. 3, 53–72 (Academic, New York, 1967).
- Totsuji, H. & Barrat, J.-L. Phys. Rev. Lett. **60**, 2484–2487 (1988).
- Diedrich, F., Peik, E., Chen, J. M., Quint, W. & Walther, H. Phys. Rev. Lett. **59**, 2931–2934 (1987).
- Blümel, R. et al. Nature **334**, 309–313 (1988).
- Wineland, D. J., Bergquist, J. C., Itano, W. M., Bollinger, J. J. & Manney, C. H. Phys. Rev. Lett. **59**, 2935–2938 (1987).
- Gilbert, S. L., Bollinger, J. J. & Wineland, D. J. Phys. Rev. Lett. **60**, 2022–2025 (1988).

ACKNOWLEDGEMENTS. We thank C. A. Schrama for discussions.

A new copper oxide superconductor containing carbon

K. Kinoshita & T. Yamada

NTT Basic Research Laboratories, Midori-cho, Musashino-shi, Tokyo 180, Japan

SUPERCONDUCTIVITY in the high-transition-temperature (high- T_c) copper oxide superconductors seems to arise from layers of copper–oxygen squares, pyramids or octahedra. Recently the compound $\text{Sr}_2\text{CuO}_2\text{CO}_3$ was found to contain layers of CuO_6 octahedra¹ (Fig. 1), suggesting that it might be made superconducting by appropriate doping. But the presence of carbonate as an impurity is known to degrade the superconducting properties of materials such as $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ and $\text{YBa}_2\text{Cu}_4\text{O}_8$ (refs 2, 3), much effort having been made to minimize residual carbon in these compounds by optimizing processing methods⁴. It is thus of some interest to see whether $\text{Sr}_2\text{CuO}_2\text{CO}_3$ can be made superconducting despite the carbonate ions incorporated in the structure. Here we report the synthesis, at 50 atm oxygen partial pressure, of superconducting $(\text{Ba}_x\text{Sr}_{1-x})_2\text{Cu}_{1+y}\text{O}_{2+2y+\delta}(\text{CO}_3)_{1-y}$ ($0.4 \leq x \leq 0.65$, $y \approx 0.1$), with T_c (onset) up to ~ 40 K and zero resistance at up to ~ 26 K. The crystal structure contains CuO_2 sheets alternating with $(\text{Ba}_x\text{Sr}_{1-x})_2\text{Cu}_y\text{O}_{2y+\delta}(\text{CO}_3)_{1-y}$ slabs, which serve as charge reservoir layers: substitution of $\sim 10\%$ copper for carbon in the slabs introduces holes into the CuO_2 sheets, making the compound superconducting.

A series of samples with nominal composition $(\text{Ba}_x\text{Sr}_{1-x})_2\text{Cu}_{1+y}\text{O}_{2+2y+\delta}(\text{CO}_3)_{1-y}$ ($0.3 \leq x \leq 1$, $0 \leq y \leq 0.5$) was synthesized from BaCO_3 , SrCO_3 , and CuO powders with purities higher than 99.9%. The powder mixtures were pressed into pellets and calcined at temperatures between 800 and 850 °C for 20 to 100 h in air. The resultant products were mixtures of $(\text{Ba}_x\text{Sr}_{1-x})_2\text{CuO}_2\text{CO}_3$, BaCuO_2 and Sr_2CuO_3 , as determined by X-ray powder diffraction and thermogravimetric and chemical

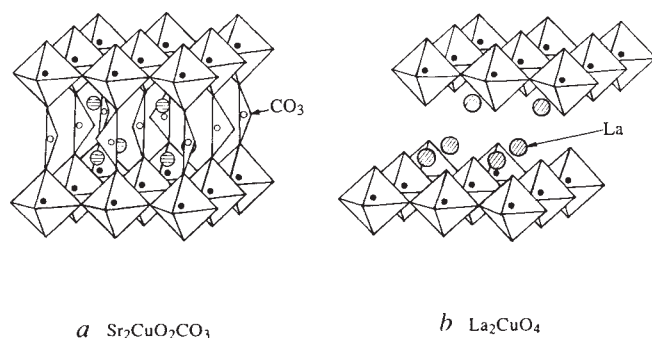


FIG. 1 Comparison of the crystal structure of $\text{Sr}_2\text{CuO}_2\text{CO}_3$ (a) and La_2CuO_4 (b). In $\text{Sr}_2\text{CuO}_2\text{CO}_3$, vertically adjacent layers of CuO_6 octahedra are connected by two oxygen atoms in the CO_3 triangles. Note that the stacking of layers of CuO_6 octahedra in $\text{Sr}_2\text{CuO}_2\text{CO}_3$ differs from that of La_2CuO_4 (ref. 11) in that the x and y coordinates of Cu atoms in two adjacent layers are the same, not shifted by $(a+b)/2$ as in La_2CuO_4 .

analyses. The time and temperature required to form $(\text{Ba}_{0.55}\text{Sr}_{0.45})_2\text{CuO}_2\text{CO}_3$ depended on the Ba/Sr ratio: longer times and higher temperatures were needed for samples richer in Ba. This is consistent with the difference in dissociation pressures between BaCO_3 and SrCO_3 (ref. 5). After calcination, a furnace for hot isostatic pressing was used to sinter the samples at 1,000 °C for 50 h in 20% O_2 plus 80% Ar at a total pressure of 250 atm (oxygen partial pressure $p_{\text{O}_2} = 50$ atm). This produced single-phase samples with compositions ranging between $(\text{Ba}_{0.65}\text{Sr}_{0.35})_2\text{Cu}_{1.1}\text{O}_{2.2+\delta}(\text{CO}_3)_{0.9}$ and $(\text{Ba}_{0.4}\text{Sr}_{0.6})_2\text{Cu}_{1.1}\text{O}_{2.2+\delta}(\text{CO}_3)_{0.9}$. The carbon content was determined by combustion. Samples with a higher Ba content contained BaCuO_2 as an impurity phase, and samples with a higher Sr content contained Sr_2CuO_3 . With insufficient calcination, BaCO_3 was also detected.

The resistivity of $(\text{Ba}_{0.55}\text{Sr}_{0.45})_2\text{Cu}_{1.1}\text{O}_{2.2+\delta}(\text{CO}_3)_{0.9}$ (measured by the standard four-probe d.c. method) decreases with decreasing temperature (Fig. 2a), showing the metallic nature of this compound's temperature dependence. The onset temperature $T_c(\text{onset})$ is ~ 40 K and $T_c(\text{zero})$ is ~ 26 K. Although $T_c(\text{onset})$ is almost independent of the Ba/Sr ratio, $T_c(\text{zero})$ is not. When the Ba content is higher than 1.2, $T_c(\text{zero})$ is lower. A Ba content of less than 0.9 also leads to a lower $T_c(\text{zero})$. The highest $T_c(\text{zero})$ was obtained for a composition around $(\text{Ba}_{0.5}\text{Sr}_{0.5})_2\text{Cu}_{1.1}\text{O}_{2.2+\delta}(\text{CO}_3)_{0.9}$.

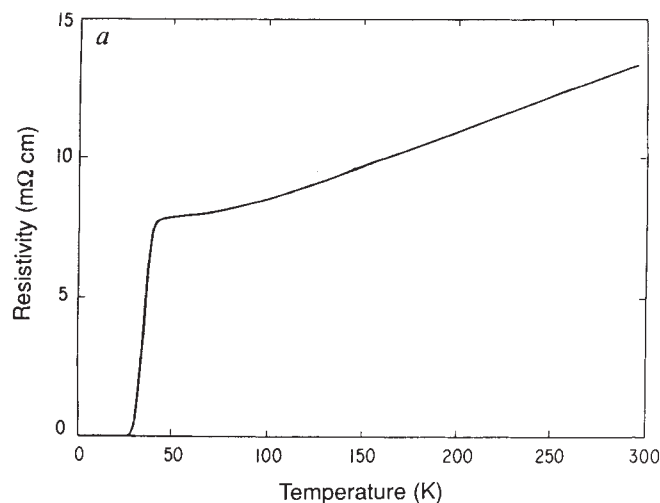


FIG. 2 a, Temperature dependence of electrical resistivity for $(\text{Ba}_{0.55}\text{Sr}_{0.45})_2\text{Cu}_{1.1}\text{O}_{2.2+\delta}(\text{CO}_3)_{0.9}$. b, Temperature dependence of magnetic susceptibility for $(\text{Ba}_{0.55}\text{Sr}_{0.45})_2\text{Cu}_{1.1}\text{O}_{2.2+\delta}(\text{CO}_3)_{0.9}$ measured in an applied field of 10 Oe under field-cooling conditions.

The bulk nature of this compound's superconductivity was confirmed by measuring the Meissner effect (Fig. 2b). Measurements were made with a SQUID magnetometer in an applied field of 10 Oe under field-cooling conditions. The $T_c(\text{onset})$ determined by measuring the magnetic susceptibility was a little lower than that determined by measuring the resistivity. The Meissner volume fraction of the sample was more than 20% that of an ideal diamagnet.

Figure 3 shows the X-ray diffraction pattern of $(\text{Ba}_{0.55}\text{Sr}_{0.45})_2\text{Cu}_{1.1}\text{O}_{2.2+\delta}(\text{CO}_3)_{0.9}$ obtained with Cu $K\alpha$ radiation. This diffraction pattern is similar to that of BaTiO_3 with the perovskite structure⁶, and it seemed possible to index the peaks on the basis of a cubic unit cell. The compound was, however, determined to be tetragonal with lattice parameters $a = 5.5639$ and $c = 7.8567$ Å (ref. 7). All peaks could be assigned as shown in Fig. 3. Low-temperature X-ray measurements showed that the crystal symmetry remained unchanged down to 20 K and showed no evidence of a structural transformation accompanying the normal-superconducting phase transition.

The crystal structure of the new superconductor has been determined by neutron powder diffraction with the auxiliary use of electron diffraction and high-resolution electron microscopy⁷. Table 1 compares the space group, lattice parameters, and Cu–O bond lengths in CuO_6 octahedra of the three compounds, $(\text{Ba}_{0.55}\text{Sr}_{0.45})_2\text{Cu}_{1.1}\text{O}_{2.2+\delta}(\text{CO}_3)_{0.9}$ (hereafter abbreviated as BSCOCO), $\text{Sr}_2\text{CuO}_2\text{CO}_3$ and La_2CuO_4 . Because of CO_3^{2-} ion disordering, BSCOCO has a unit cell different from that of $\text{Sr}_2\text{CuO}_2\text{CO}_3$. The unit cell of BSCOCO is also different from that of La_2CuO_4 because there is no $(a+b)/2$ shift between vertically adjacent CuO_6 layers. Therefore, for $\text{Sr}_2\text{CuO}_2\text{CO}_3$ and La_2CuO_4 , we list in Table 1 the lattice parameters of a pseudo unit cell reduced to the same unit as BSCOCO, where $a' = a/\sqrt{2}$, and $c' = c/2$. Only partial ordering of Ba and Sr atoms has been observed between the two alkaline earth element sites in the superconductor, but comparison of the lattice parameters of BSCOCO and $\text{Sr}_2\text{CuO}_2\text{CO}_3$ shows that substitution of larger Ba^{2+} ions for the smaller Sr^{2+} ions increases the lattice spacing in the [001] direction more than in the [100] direction. This increases the bond lengths between Cu and apical-site oxygen in CuO_6 octahedra rather than those between Cu and plane-site oxygen. As a result, bond lengths of Cu–O(apex) in the new superconductor are much longer than in $\text{Sr}_2\text{CuO}_2\text{CO}_3$ and La_2CuO_4 , and also much longer than the value of ~ 2.3 Å between Cu(2) and apical-site oxygen O(4) in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$, whereas the bond lengths between Cu and plane-site oxygen

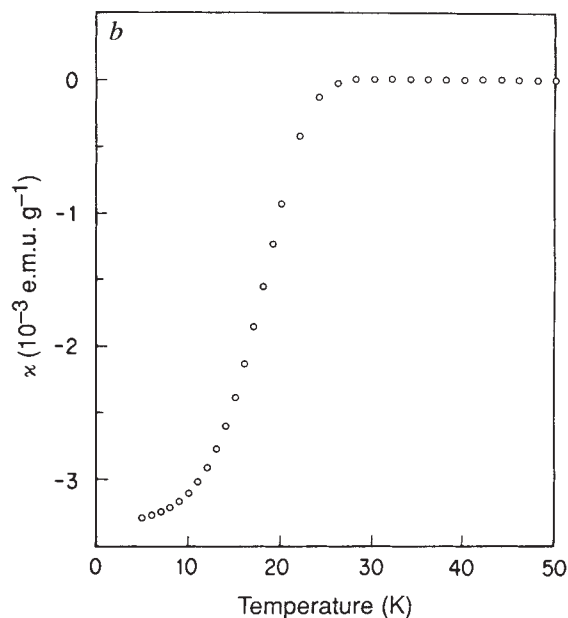
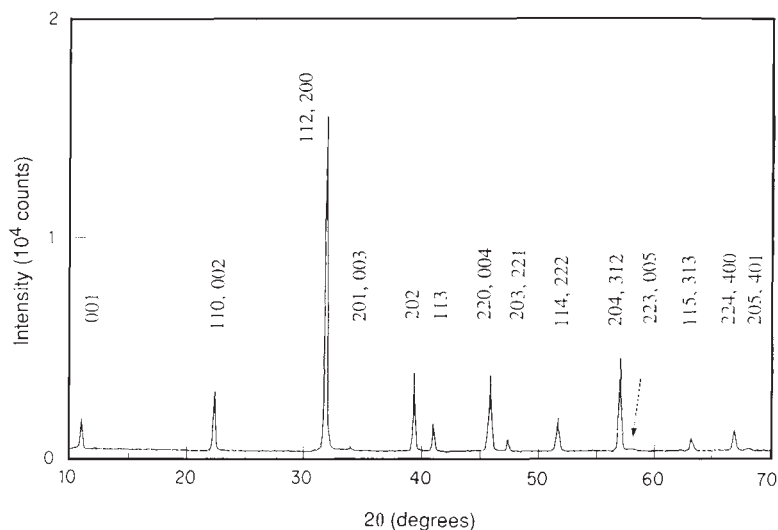


FIG. 3 X-ray powder diffraction pattern of $(\text{Ba}_{0.55}\text{Sr}_{0.45})_2\text{Cu}_{1.1}\text{O}_{2.2+\delta}(\text{CO}_3)_{0.9}$ for Cu K α radiation. The lattice parameters of the tetragonal structure are $a=5.5639$ Å and $c=7.8567$ Å.



remain at the values typical of oxide superconductors⁸. Such longer Cu–O(apex) bond lengths imply that the bonds between Cu and apical-site oxygen are very weak, and the crystal structure of the new superconductor can be regarded as alternate stacking of CuO_2 sheets and $(\text{Ba}_x\text{Sr}_{1-x})_2\text{Cu}_y\text{O}_{2y+\delta}(\text{CO}_3)_{1-y}$ carbonate slabs rather than as having layers of CuO_6 octahedra.

In the superconductor, ~10% of the C atoms in the carbonate slabs are replaced by Cu atoms. The longer c -spacing caused by the substitution of Ba for Sr may facilitate such replacement, because the shortest Cu–O bond length of ~1.9 Å in the superconductor is much longer than C–O bond length of ~1.3 Å. Crystal structure analysis⁷ shows that Cu atoms replace carbon sites randomly, but the oxygen arrangements around these Cu atoms are very similar to those around chain site Cu(1) atoms in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$. Therefore, copper atoms in the carbonate slabs have excess oxygen around them and may act to introduce hole carriers into the CuO_2 sheets, as do those in the linear chains in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ (ref. 9). Iodometric titration showed that the average valence of Cu was ~2.20 for superconductors and ~2.05 for nonsuperconducting mixtures of $(\text{Ba}_x\text{Sr}_{1-x})_2\text{CuO}_2\text{CO}_3$, BaCuO_2 and Sr_2CuO_3 calcined at 850 °C, supporting our idea about the role of substituted Cu atoms and the excess oxygen around them. If the valences of Ba, Sr and C are taken as +2, +2 and +4, respectively, the amount of excess oxygen δ is calculated to be ~0.1 for the superconductor. The role of high oxygen pressure during the synthesis of the superconductor is to prevent the dissociation of carbonate ions at high temperatures (~1,000 °C) and to supply excess oxygen around substituted Cu atoms in the carbonate slabs. If the $(\text{Ba}_x\text{Sr}_{1-x})_2\text{Cu}_y\text{O}_{2y+\delta}(\text{CO}_3)_{1-y}$ layer can be regarded as a block

layer¹⁰, $(\text{Ba}_x\text{Sr}_{1-x})_2\text{Cu}_{1+y}\text{O}_{2+2y+\delta}(\text{CO}_3)_{1-y}$ might be classified as having a new type of block layer. □

Received 4 March; accepted 21 April 1992.

1. Miyazaki, Y. *et al.* *Physica C* **191**, 434–440 (1992).
2. Shaw, T. M., *et al.* *J. Mater. Res.* **5**, 1176–1184 (1990).
3. Zandbergen, H. W., Cava, R. J., Krajewski, J. J. & Peck, Jr, W. F. *Physica C* **179**, 227–242 (1991).
4. Lindemer, T. B., Washburn, F. A., MacDougall, C. S. & Cavin, O. B. *Physica C* **174**, 135–143 (1991).
5. Lander, J. J. *J. Am. chem. Soc.* **73**, 5794–5797 (1951).
6. Naka, S., Nakakita, F., Suwa, Y. & Inagaki, M. *Bull. chem. Soc. Jpn* **47**, 1168–1171 (1974).
7. Izumi, F. *et al.* *Physica C* (in the press).
8. David, W. I. F. *et al.* *Nature* **327**, 310–312 (1987).
9. Tokura, Y., Torrance, J. B., Huang, T. C. & Nazzari, A. I. *Phys. Rev.* **B38**, 7156–7159 (1988).
10. Tokura, Y. & Arima, T. *Jpn J. appl. Phys.* **29**, 2388–2402 (1990).
11. Grande, V. B., Müller-Buschbaum, H. & Schweizer, M., *Z. Anorg. allg. Chem.* **428**, 120–124 (1977).

ACKNOWLEDGEMENTS. We thank T. Kimura for his support and encouragement.

A sequential-decision strategy for abating climate change

James K. Hammitt*, Robert J. Lempert* & Michael E. Schlesinger†

* RAND, 1700 Main Street, Santa Monica, California 90407-2138, USA

† Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign, 105 South Gregory Avenue, Urbana, Illinois 61801, USA

CURRENT debate on policies for limiting climate change due to greenhouse-gas emissions focuses on whether to take action now or later, and on how stringent any emissions reductions should be in the near and long term. Any reductions policies implemented now will need to be revised later as scientific understanding of climate change improves. Here we consider the effects of a sequential-decision strategy (Fig. 1) consisting of a near-term period (1992–2002) during which either moderate emissions reductions (achieved by energy conservation only) or aggressive reductions (energy conservation coupled with switching to other fuel sources) are begun, and a subsequent long-term period during which a least-cost abatement policy is followed to limit global mean temperature change to an optimal target ΔT^* . For each policy we calculate the global mean surface temperature change $\Delta T(t)$ using a simple climate/ocean model for climate sensitivities ΔT_{2x} (the response to doubled CO_2 concentrations) of 4.5, 2.5, 1.5 and 0.5 °C. The policy beginning with moderate reductions is less expensive than that with aggressive reductions if $\Delta T^* > 2.9, 2.1, 1.5$ and 0.9 °C respectively; otherwise, the aggressive-reductions policy is cheaper. We suggest that this approach should assist in choosing

TABLE 1 Space group, lattice parameters and Cu–O bond lengths in CuO_6 octahedra

	BSCOCO* (ref. 7)	$\text{Sr}_2\text{CuO}_2\text{CO}_3$ (ref. 1)	La_2CuO_4 (ref. 11)
Space group	$P4_2/m$ or $P4_2/m$	$I\bar{4}$	$Cmca$
Lattice parameters† (Å)	$a=5.5639$ $c=7.8567$	$a'=5.519$ $c'=7.4965$	$a=5.406$ $b=5.370$ $c'=6.575$
Bond lengths (Å)			
Cu–O(plane)	1.952 (2×) 1.982 (2×)	1.952 (2×) 1.957 (2×)	1.907 (4×)
Cu–O(apex)	2.604 3.059	2.559 2.820	2.459 (2×)

* Abbreviation of $(\text{Ba}_{0.55}\text{Sr}_{0.45})_2\text{Cu}_{1.1}\text{O}_{2.2+\delta}(\text{CO}_3)_{0.9}$.

† In $\text{Sr}_2\text{CuO}_2\text{CO}_3$ and La_2CuO_4 , the lattice parameters of the reduced unit-cell are listed, where $a'=a/\sqrt{2}$, and $c'=c/2$.

realistic targets and in determining how best to implement emissions reductions in the short and long term.

We characterize scientific uncertainties by the sensitivity ΔT_{2x} and the climate target^{1,2} ΔT^* (or maximum realized value of $\Delta T(t)$), implicitly chosen so that the present value (1992–2100) of the marginal costs of abatement equals that of damages after accounting for adaptation³. For the policies considered here, we assume that global damages depend only on ΔT^* , although actual damages may depend on the rate, magnitude and duration of regional climatic changes.

Our simulation of anthropogenic greenhouse-gas (GHG) emissions and the costs of their reduction uses a heuristic model based on the link between energy use and GHG releases. For simplicity, we measure GHGs in units of equivalent carbon dioxide⁴ (ECD). For 1765–1990 we use the ECD estimated by the Intergovernmental Panel on Climate Change (IPCC)⁴. For year $t \geq 1990$, the global flux of ECD is $F(t) = F_0 B(t) I(t) E(t)$, where $F_0 = 11.0$ gigatonnes (Gt) carbon is the global 1990 ECD emission and $B(t) = 1 + 0.0364(t - 1990)$ is a base-case trajectory chosen so that $F(t)$ reproduces the IPCC 1990–2100 base case⁴. $I(t)$ and $E(t)$ represent policy-induced shifts in primary energy consumption per unit economic product and in GHG emissions per unit energy, respectively. For $t < 1992$ and for all t in the base case, $I(t) = E(t) = 1$.

ECD emissions can be reduced by various means, which vary in cost-effectiveness and speed of implementation. We consider two options, 'energy conservation' and 'fuel switching'.

Energy conservation represents a set of low-cost, quickly implemented policies; if adopted at t_0 , energy intensity $I(t)$ falls 30% over 20 years. For $t \geq t_0$,

$$I(t) = (1 - \gamma) + [\gamma / (1 - \varepsilon)] [1 + e^{\rho(t - t_0 - r)}]^{-1} \quad (1)$$

where $\gamma = 0.3$ and the transition half-life $r = 10$ yr; to satisfy $I(t_0) = 1$, $\rho = -(1/r) \ln [\varepsilon / (1 - \varepsilon)]$, and we choose $\varepsilon = 0.01$ to obtain reasonable slopes⁵ at $t_0 + r$. The logistic decline is characteristic of technological diffusion^{5,6}.

Fuel switching represents a set of high-cost, slowly implemented emission-reduction technologies. Its cost and effectiveness are simulated by assuming that all emissions are produced by long-lived capital equipment using either emitting (for example fossil fuel) or non-emitting (for example nuclear, solar, biomass) technologies; for both technologies, the construction period is 10 yr and the operating period is 30 yr. In the base

case, only emitting technologies are considered. The policy decisions are whether to begin substituting non-emitting for emitting equipment, and if so, the rate at which to substitute, specified as the transition half-life, r , or the time until half the capital stock then operating is non-emitting.

We consider two near-term policies, aggressive and moderate reduction. Under the aggressive-reduction policy, both conservation and fuel-switching options are adopted beginning in 1992, the latter with a 40-yr transition half-life. Under the moderate-reduction policy, only the conservation option is taken in the first period. At the beginning of the second period, 2002, the values of ΔT_{2x} and ΔT^* are assumed to become known, and the rate of fuel switching is chosen so that $\Delta T(t)$ peaks at the revealed target ΔT^* given the revealed sensitivity ΔT_{2x} . For $t \geq t_0$, $E(t)$ is given by equation (1) with $\gamma = 1$. Because of the 10-yr construction period, initiation of fuel switching in 2002 cannot affect $E(t)$ until a decade later. Consequently, for the moderate policy, $t_0 = 2012$ and r is chosen to yield the desired ΔT^* . For the aggressive policy, for $2002 \leq t \leq 2012$, $t_0 = 2002$ and $r = 40$ yr. For $t > 2012$, r is chosen to yield the desired ΔT^* and $t_0 = 2,012 - r/4$ so that $E(t)$ is continuous at $t = 2012$.

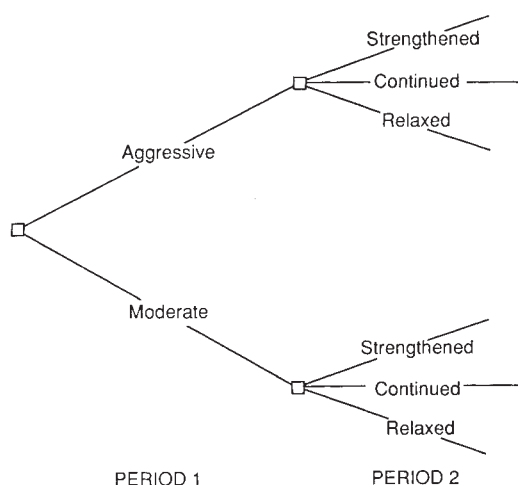


FIG. 1 Decision tree for sequential policy choice. At the beginning of the first period (1992), either the moderate reduction or aggressive reduction near-term abatement policy is selected. At the beginning of the second period (2002), uncertainties about climate sensitivity and the damages of climate change are substantially resolved and the chosen policy is amended to limit climate change to the 'optimal' level.

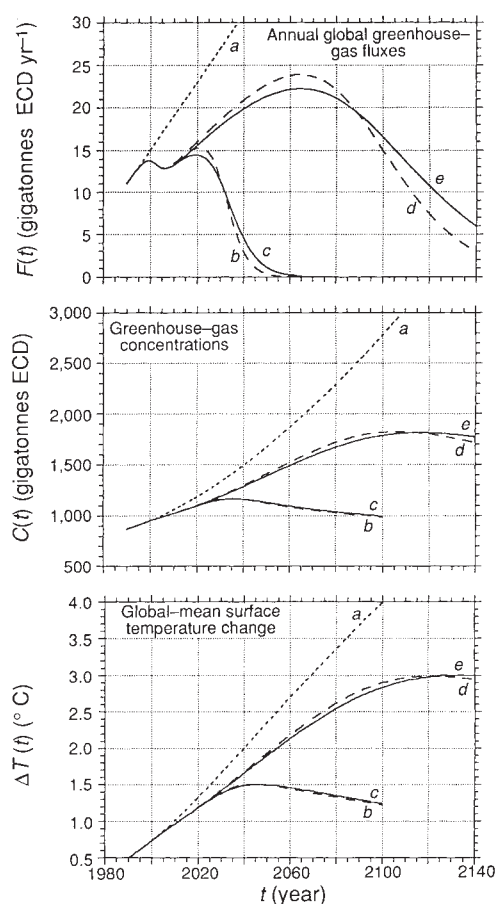


FIG. 2 Annual global greenhouse-gas fluxes $F(t)$ (top panel), concentrations $C(t)$ (middle panel) and global-mean surface temperature change $\Delta T(t)$ (lower panel) for $\Delta T_{2x} = 2.5^\circ\text{C}$, for base case and for moderate and aggressive policies for $\Delta T^* = 1.5$ and 3°C . Under the moderate-reduction policy, conservation is implemented beginning in 1992, and fuel switching with a 20-yr half-life (for $\Delta T^* = 1.5^\circ\text{C}$) or 80-yr half-life (for $\Delta T^* = 3^\circ\text{C}$) begins in 2002. Under the aggressive-reduction policy, conservation and fuel switching with a 40-yr half-life are implemented beginning in 1992 and fuel switching is changed to a 27-yr half-life (for $\Delta T^* = 1.5^\circ\text{C}$) or 108-yr half-life (for $\Delta T^* = 3^\circ\text{C}$) in 2002. We compare the costs of policies that reach the same peak $\Delta T(t) = \Delta T^*$. a, Base case. b, Moderate-reduction policy, $\Delta T^* = 1.5^\circ\text{C}$. c, Aggressive-reduction policy, $\Delta T^* = 1.5^\circ\text{C}$. d, Moderate-reduction policy, $\Delta T^* = 3^\circ\text{C}$. e, Aggressive-reduction policy, $\Delta T^* = 3^\circ\text{C}$.

FIG. 3 Annual costs (1992–2100, 5% discount rate) of aggressive- and moderate-reduction policies as a function of ΔT^* and ΔT_{2x} . The dashed line demarcates the region where the moderate-reduction policy is less expensive, which corresponds to fuel-switching half-lives in the second period of ~40–50 yr for the moderate-reduction and 50–70 yr for the aggressive-reduction policy. Depending on ΔT_{2x} , certain values of ΔT^* cannot be achieved because of the 'warming commitment' due to previous GHG emissions and those that will occur before existing capital stock can be replaced. *a*, Moderate reductions and $\Delta T_{2x} = 0.5^\circ\text{C}$. *b*, Aggressive reductions and $\Delta T_{2x} = 0.5^\circ\text{C}$. *c*, Moderate reduction and $\Delta T_{2x} = 1.5^\circ\text{C}$. *d*, Aggressive reductions and $\Delta T_{2x} = 1.5^\circ\text{C}$. *e*, Moderate reductions and $\Delta T_{2x} = 2.5^\circ\text{C}$. *f*, Aggressive reductions and $\Delta T_{2x} = 2.5^\circ\text{C}$. *g*, Moderate reductions and $\Delta T_{2x} = 4.5^\circ\text{C}$. *h*, Aggressive reductions and $\Delta T_{2x} = 4.5^\circ\text{C}$.

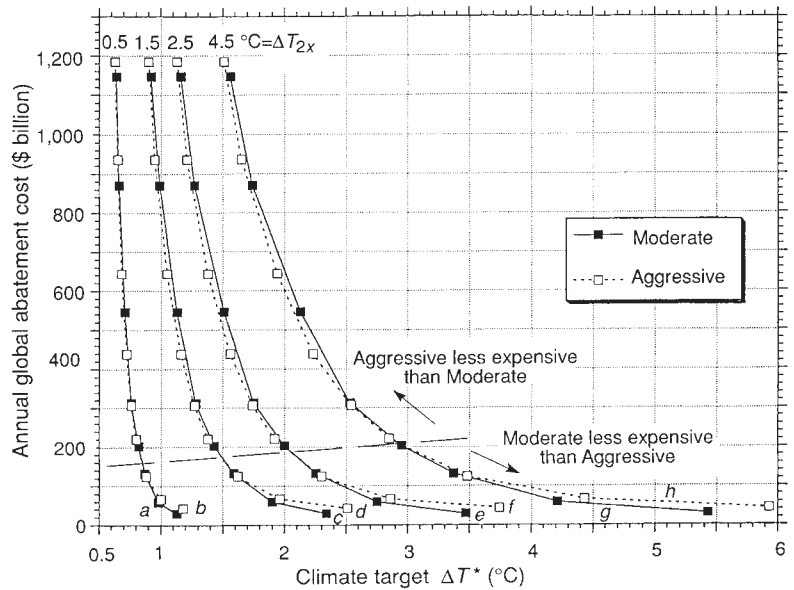


Figure 2 illustrates emissions under the two policies for $\Delta T_{2x} = 2.5^\circ\text{C}$, and $\Delta T^* = 1.5$ and 3°C .

Atmosphere ECD concentration $C(t)$ is modelled as a function of anthropogenic emissions

$$C(t) = C_0 + \int_{-\infty}^t F(\tau) \left\{ \alpha_1 + \sum_{i=2}^5 \alpha_i \exp[(\tau - t)/\lambda_i] \right\} d\tau \quad (2)$$

where $C_0 = 588$ Gt carbon is the pre-industrial ECD concentration, $\alpha_{1-5} = 0.13, 0.20, 0.32, 0.25, 0.10$ and $\lambda_{2-5} = 363, 74, 17, 2$ yr. Equation (2) is a linear impulse-response function for CO_2 , parameterized to include the effects of the primary ocean sinks, but not the biota or any modification of the carbon cycle accompanying climate change⁷. For each ECD-concentration scenario, $\Delta T(t)$ is simulated using the energy-balance climate/upwelling-diffusion ocean model of Schlesinger and Jiang⁸ for $\Delta T_{2x} = 4.5, 2.5, 1.5$, as chosen by IPCC⁴, and $\Delta T_{2x} = 0.5^\circ\text{C}$, as proposed by Lindzen⁹. The temperature trajectory for each ΔT_{2x} is calculated beginning from $\Delta T(1765) = 0$, but is then offset so that $\Delta T(1990) = 0.5^\circ\text{C}$, consistent with observations^{4,10,11}. Illustrations of $C(t)$ and $\Delta T(t)$ are shown in Fig. 2. As $F(t)$ goes to zero, both $C(t)$ and $\Delta T(t)$ peak and begin to fall. These declines are sensitive to the modelling of ECD concentrations (Equation 2); if realistic, such behaviour could be important in choosing long-term emission levels.

Consistent with recent claims that substantial efficiency improvements are available at low or even negative cost^{12–14}, the incremental cost of the conservation option is \$5 per tonne of avoided carbon emissions. The incremental cost (in billions (10^9) of 1990 dollars) due to fuel switching is the difference between policy-induced and base-case expenditures on capital equipment. Fuel switching is divided between low-cost and high-cost components, reflecting increasing marginal abatement costs. For each emission scenario, annual costs $K(t)$ are given by

$$K(t) = \sum_{j=0}^2 \left[\sum_{i=-9}^0 \kappa_j n_{ji}(t) + \sum_{i=1}^{30} \sigma_j n_{ji}(t) \right] \quad (3)$$

where $n_{0i}(t)$, $n_{1i}(t)$ and $n_{2i}(t)$ are, respectively, the energy used by emitting and low-cost and high-cost non-emitting equipment of age i in year t ($-9 \leq i \leq 0$ indexes equipment in construction), σ_j are annual operating costs and $10\kappa_j$ are total capital costs, allocated uniformly over the construction period. Capital and annual operating costs for emitting equipment, based on coal-fired electric plants, are $\kappa_0 = \$0.015$ per kWh ($\$1 \text{ W}^{-1}$ installed capacity, 75% capacity factor) and $\sigma_0 = \$0.025$ per kWh (ref.

15). For non-emitting equipment these are increased proportionally ($\kappa_1 = \$0.017$ per kWh; $\sigma_1 = \$0.028$ per kWh; $\kappa_2 = \$0.022$ per kWh; $\sigma_2 = \$0.037$ per kWh) to yield lifetime costs of \$50 per tonne and \$200 per tonne avoided carbon emissions^{16,17}.

We assume that future demand for energy services is known, that sufficient capital equipment enters construction each year to satisfy increased demand and to replace equipment that will retire a decade later, and that sufficient non-emitting equipment is installed to satisfy the emissions constraint. Consequently, $n_{j-9}(t)$ are chosen for $t \geq 1990$ so that capital equipment operating a decade hence satisfies the demand constraint

$$\sum_{i=-9}^{20} \sum_{j=0}^2 n_{ji}(t) = Q_0 B(t+10) I(t+10) \quad (4)$$

the emission constraint

$$\sum_{i=-9}^{20} n_{0i}(t) = Q_0 B(t+10) I(t+10) E(t+10) \quad (5)$$

and the constraint on applicability of low-cost non-emitting capital

$$\sum_{i=-9}^{20} n_{1i}(t) \leq \frac{1}{2} Q_0 B(t+10) I(t+10) \quad (6)$$

For $-8 \leq i \leq 30$ and all j , $n_{ji}(t+1) = n_{j,i-1}(t)$. For $t < 1990$ and all i , $n_{1i}(t) = n_{2i}(t) = 0$ and $n_{0i}(t)$ are chosen to reproduce pre-1990 ECD fluxes. The magnitude of the energy-using sector, Q_0 , is determined by dividing 1990 world ECD emissions, $F_0 = 11.0$ Gt carbon, by the average ratio of world industrial CO_2 emissions to commercial energy use, 7.2 tonne C per 10^5 kWh. Unless equation (6) binds, $n_{2,-9}(t) = 0$. If $n_{0,-9}(t) = 0$ is insufficient to satisfy equation (5), emitting equipment of ages $i \leq 20$ is scheduled for early retirement, oldest first.

The incremental 1992–2100 costs, annualized at 5% discount rate, are presented in Fig. 3. With discounting, the 2100 end-date has little effect on either total costs or the cost difference between policies producing the same ΔT^* . Present-value abatement costs are smaller for the moderate policy than for the aggressive policy if and only if ΔT^* exceeds a ΔT_{2x} -dependent critical value: 2.9, 2.1, 1.5 and 0.9°C , for $\Delta T_{2x} = 4.5, 2.5, 1.5$ and 0.5°C , respectively. Abatement costs are more sensitive to the climate target and climate sensitivity than to the choice between the two policies, with higher abatement costs for smaller climate targets and larger climate sensitivities.

The critical values of ΔT^* are independent of conservation

costs and the level of non-emitting capital costs, and are modestly sensitive (shift $\leq 30\%$) to the cost ratio of high-cost to low-cost non-emitting capital (1–10), the duration of the near-term period (5–20 yr), technological innovation that reduces the incremental cost of non-emitting equipment by 2% per yr, and a 10% discount rate (a 0% discount rate increases the critical values 40–60%). Total abatement costs depend on fuel-switching costs, technological innovation and the discount rate and, to a lesser extent, on the duration of the first period and the cost of conservation.

The apparent desirability of postponing aggressive abatement options for ΔT^* greater than a ΔT_{2x} -dependent critical value depends on the assumption that nonlinear responses in the climate system will not preclude offsetting additional GHGs emitted in the next 10 yr with faster emission reductions thereafter (Fig. 2), and on the feasibility of rapid global emission

reductions. For $\Delta T_{2x} = 2.5^\circ\text{C}$, a 10-yr delay in beginning fuel switching can be overcome by an apparently modest decrease in the transition half-life from 108 to 80 yr for $\Delta T^* = 3^\circ\text{C}$, but requires a sharper decrease from 27 to 20 yr for $\Delta T^* = 1.5^\circ\text{C}$. The ~ 100 -yr half-life is roughly the rate at which oil and gas were adopted worldwide⁵. The ~ 25 -yr half-life is comparable to the French nuclear-power programme¹⁸, but such a rapid transition would be difficult to achieve globally, and construction bottlenecks (not simulated) could further increase costs. Schlesinger and Jiang⁸ also conclude that a 10-yr delay may have little effect on the ensuing climate trajectory. This suggests that the rate at which emissions can be reduced may be more important than when reductions begin, and that attention should be directed towards identifying factors that influence the rate of technological diffusion worldwide, and developing expertise and institutions to accelerate the transition. $\square$

Received 19 December 1991; accepted 5 May 1992.

1. Rijsberman, F. R. & Swart, R. J. *Targets and Indicators of Climatic Change* (Stockholm Environment Institute, 1990).
2. Wirth, D. A. & Lashof, D. A. *Ambio* **19**, 305–310 (1990).
3. Hammit, J. K. & Lempert, R. J. in *Proc. Conf. Global Climate Change: Los Alamos National Laboratory, October 21–24, 1991* (eds Rosen, L. & Glasser, R.) (Am. Inst. Phys., in the press).
4. Houghton, J. T., Jenkins, G. J. & Ephraums, J. J. (eds) *Climate Change: The IPCC Scientific Assessment* (Cambridge Univ. Press, 1990).
5. Hafele, W. (ed.) *Energy in a Finite World: A Global Systems Analysis* (Ballinger, Cambridge, Massachusetts, 1981).
6. Fisher, J. C. & Pry, R. H. *Simple Substitution Model of Technological Change, General Electric Co. Res. Dev. Center Rep. 70-C-215* (Schenectady, New York, 1970).
7. Maier-Reimer, E. & Hasselmann, K. *Clim. Dynam.* **2**, 63–90 (1987).
8. Schlesinger, M. E. & Jiang, X. *Nature* **350**, 219–221 (1991).
9. Lindzen, R. S. *Bull. Am. met. Soc.* **71**, 288–299 (1990).
10. Hansen, J. & Lebedeff, S. *J. geophys. Res.* **92**, 13345–13372 (1987).

11. Jones, P. D., Wigley, T. M. L. & Wright, P. B. *Nature* **322**, 430–434 (1986).
12. National Academy of Sciences *Policy Implications of Greenhouse Warming* (National Academy, Washington, DC, 1991).
13. Office of Technology Assessment, US Congress *Changing by Degrees: Steps to Reduce Greenhouse Gases*, OTA-D-482 (Washington, DC, 1991).
14. Fickett, A. P., Gellings, C. W. & Lovins, A. B. *Scient. Am.* **263**, 64–74 (1990).
15. Office of Technology Assessment, US Congress *New Electric Power Technologies*, OTA-E-246 (Washington, DC, 1985).
16. Manne, A. S. & Richels, R. G. *Energy J.* **12**, 87–108 (1991).
17. Nordhaus, W. D. *Energy J.* **12**, 37–65 (1991).
18. Ausubel, J. H. *Nature* **350**, 649–652 (1991).

ACKNOWLEDGEMENTS. We thank J. Adams for assistance with modelling and calculating and A. Ghanem for assistance in running the simple climate/ocean model. This research was supported by RAND, the US NSF and the US Department of Energy, Carbon Dioxide Research Program, Office of Health and Environmental Research.

Sequestration of CO₂ in the deep ocean by shallow injection

Peter M. Haugan & Helge Drange

Nansen Environmental and Remote Sensing Center, Edvard Griegs vei 3A, N-5037 Solheimsvik, Norway

To alleviate the increasing atmospheric concentrations of carbon dioxide, the principal greenhouse gas, Marchetti¹ proposed that CO₂ might be separated from flue gases and injected into the oceans. This, and subsequent studies^{2–5}, emphasized the need to inject the gas at great depths or in sinking currents to avoid rapid outgassing to the atmosphere. Here we show that, to the contrary, the increase in water density that results from CO₂ dissolution may be sufficient to transport the dissolved gas to lower depths even for shallow injection (in the upper 200–400 m of the ocean). If the CO₂ is injected near the shore, gravity currents will carry the dense, CO₂-rich waters along the bottom slope towards deep water. Shallow injection near the shore will be less expensive in terms of energy and capital than deep-ocean injection. We suggest that the coast of Norway, in the vicinity of existing oil and gas fields and of planned gas power plants, provides an example of a region where the negative buoyancy of CO₂-enriched sea water would transport the gas from emission sites to the deep ocean. The effect of such measures on marine life downstream of the injection point remains to be evaluated.

Sinking currents that would transport and spread injected CO₂ into large water masses, such as the outflow from the Mediterranean to the Atlantic, have previously been identified¹ as suitable injection sites. Experiments with a global carbon-cycle model based on a three-dimensional ocean-circulation model have confirmed the viability of the Marchetti scenario, even for injection of as much as one third of the total expected global emissions in the next century into the North Atlantic near Spain⁶. Similar model studies injecting much smaller amounts of CO₂ from a 1 GW coal-fired power plant into the

North Pacific where the currents are not as favourable, indicated the need for deep injection ($> 1,000$ m) to avoid rapid outgassing of CO₂ to the atmosphere^{4,5}.

The perturbations in total dissolved inorganic carbon concentration (C_T ; mol m⁻³) in the ocean due to injection will be small when the injected carbon is diluted in large water masses, such as in the large-scale model experiments. But carbon concentrations may increase significantly near an injection site. For C_T greater than 3 mol m⁻³, that is about 50% above typical natural levels in the ocean, variations in C_T are well approximated by variations in CO₂ + H₂CO₃ (Fig. 1). The change in seawater density ρ_w (kg m⁻³) due to change in C_T (mol m⁻³) may then be expressed as:

$$d\rho_w/dC_T = [M_{\text{CO}_2} - \rho_0 \times V] + (\partial\rho_w/\partial T)H/c_p, \quad (1)$$

where M_{CO_2} is the molar mass of CO₂; $M_{\text{CO}_2} = 44.01 \times 10^{-3}$ kg mol⁻¹, ρ_0 is the ambient density of sea water, V is the partial molar volume of CO₂ in sea water (m³ mol⁻¹), T is temperature, H is the heat of solution of CO₂ in sea water (J mol⁻¹) and c_p (J m⁻³ K⁻¹) is the heat capacity of sea water. With conservative (high) values $V = 34 \times 10^{-6}$ m³ mol⁻¹ (ref. 7) and $\rho_0 = 1.03 \times 10^3$ kg m⁻³, the term in brackets in equation (1) becomes 9.0×10^{-3} kg mol⁻¹. The last term in equation (1) describes the thermal expansion of sea water due to released heat of solution of CO₂. At low pressures, $H \approx 24 \times 10^3$ J mol⁻¹ (ref. 8), and with $c_p = 4.0 \times 10^6$ J m⁻³ K⁻¹, the temperature increase becomes 6×10^{-3} K for each mol m⁻³ of C_T added. The corresponding density decrease ($\partial\rho_w/\partial T$ is negative) is negligible at very low temperatures but becomes about 1×10^{-3} kg mol⁻¹ at temperatures around 10°C and approaches 2×10^{-3} kg mol⁻¹ at temperatures above 25°C because of the strong increase of $|\partial\rho_w/\partial T|$ with temperature⁹. A conservative estimate of the right-hand side of equation 1 is 8×10^{-3} kg mol⁻¹.

The dissolvable amount of carbon at 200–400 m depth is about 10^3 mol m⁻³ (Fig. 2). If such C_T is obtained by injection, the net seawater density increase will be about 8 kg m⁻³, which is 6–7 times the density difference between the Mediterranean outflow water and its surroundings¹⁰. At $C_T \approx 10$ mol m⁻³, cor-

TABLE 1 Ascent (m) of CO₂ bubbles until dissolution

Release depth (m)	Degree of dissolution (%)	Initial bubble radius (cm)			Background concentration C_∞ (initial bubble radius=1 cm)		
		0.5	1	3	$0.2 \times C_s$	$0.4 \times C_s$	$0.8 \times C_s$
200	90	4	12	85	16	21	71
	99	5	16	111	20	28	93
300	90	5	16	109	21	28	90
	99	7	21	139	27	37	117

Ascent of CO₂ bubbles until near complete dissolution for different release depths, initial equivalent radii and background carbon concentration in sea water C_∞ . With the pressure in the bubble given by the hydrostatic pressure in the water column and C_b by an appropriate equation of state for the CO₂ gas (we use a van der Waals equation, adjusted to fit the Benedict-Webb-Rubin equation²¹ to within 1.5% for $0 \leq P \leq 45$ bar and $5^\circ\text{C} \leq T \leq 20^\circ\text{C}$), we have solved equation (2) numerically for $r(t)$ using a fourth-order Runge-Kutta scheme²². Basic water temperature is 7°C and salinity is 35‰. For the cases with elevated C_∞ , the temperature increase in the water because of heat of solution (see text) has been taken into account in C_s (and C_∞); it increases the vertical ascent by about 10% in the $C_\infty=0.8C_s$ case.

responding to an increase of the present level by a factor of 4–5, the excess density is about 0.1 kg m^{-3} . Density differences of this magnitude are sufficient for development of negatively buoyant plumes on kilometre scales in weakly stratified environments¹¹.

Pumping of CO₂-enriched sea water to a depth of 1,000 m and release into a stagnant, neutrally stable ocean gives rapid vertical sinking for realistic discharge rates². Artificial enrichment at elevated pressure and subsequent injection of enriched water would also be possible at shallower depths. Another alternative would be to inject gas directly into the ocean. For injection of gaseous CO₂ the negative buoyancy of enriched sea water will depend on the speed of uptake of CO₂ from ascending bubbles. The evolution of a CO₂ bubble is governed by the mass transfer equation:

$$d/dt(4/3\pi r^3 C_b) = -4\pi r^2 k(C_s - C_\infty) \quad (2)$$

where r (in m) is the equivalent radius of the bubble; C_b (mol m^{-3} gas) is the concentration of CO₂ gas in the bubble; C_s (mol m^{-3} sea water) is the saturation concentration of dissolved CO₂ at the surface of the bubble (given in Fig. 2), and C_∞ (mol m^{-3}) is the concentration of dissolved CO₂ in the sea water away from the bubble surface. The mass transfer coefficient k (m s^{-1}) based on boundary layer theory is taken as¹²:

$$k = [(4/\pi)(1 - 2.84/\sqrt{\text{Re}})(Du/2r)]^{1/2} \quad (3)$$

where D ($\text{m}^2 \text{s}^{-1}$) is the molecular diffusion coefficient for CO₂ in sea water¹³, u (m s^{-1}) is the vertical bubble velocity and Re is the Reynolds number ($\text{Re} = 2rup_w/\mu_w$, where μ_w is the viscosity of sea water). Experimentally, the mass transfer from carbon dioxide bubbles in water follows equation (3) closely for $r > 0.5$ cm (ref. 14). The vertical velocity of the bubble is estimated from computed and measured terminal velocities for bubbles in slightly contaminated water¹²; it increases from about 30 cm s^{-1} for bubbles with $r = 1$ cm to about 50 cm s^{-1} for bubbles with $r = 3$ cm.

Results of numerical solution of equation (2) (Table 1) indicate near complete dissolution of small bubbles after 5–40 m ascent, even when they are released into water which already contains considerable amounts of carbon. The $r = 3$ cm case may be taken as a conservative upper limit; interpretation of observations of bubble breakup for various fluid systems¹⁵ indicate that CO₂ bubbles of about this size in sea water may rapidly disintegrate into smaller bubbles. In practical applications, initial bubble size may be controlled by a mechanical diffuser device. Our calculations show that even if small bubbles merge into larger ones or if no diffuser device is used, bubble dissolution will be rapid.

A 1 GW gas power plant would emit about 3×10^9 kg CO₂ per yr (Norwegian State Pollution Authority, personal communication), which corresponds to 2.2×10^3 mol s^{-1} . The minimum sea

water flux to dissolve this amount at a pressure of 20 bar is about $2.5 \text{ m}^3 \text{s}^{-1}$. A significant buoyancy effect would be obtained even if the injected CO₂ is transported out of the bubble zone with seawater flux as high as $200 \text{ m}^3 \text{s}^{-1}$. On the basis of estimated bubble zone height, it seems that bubble zone sea water renewal rates between these limits can be expected. This indicates that injection of CO₂ in the gaseous phase and *in situ* dissolution may be an alternative to artificial dissolution at elevated pressure followed by pumping of enriched water to the injection depth.

If the injection point is in shallow water with weak ambient stratification, enriched water will reach the bottom where it will spread and may fill up topographic depressions or, in the case of a sloping bottom, develop into a bottom gravity current towards deep water^{10,16}. The final destination of bottom gravity currents may be calculated with some confidence from ambient stratification, bottom friction parameters and an entrainment hypothesis^{17,18}, given the source of (negative) buoyancy. Obviously, a route towards the deep ocean and ultimate dilution in large water masses is to be preferred over disposal in shallow

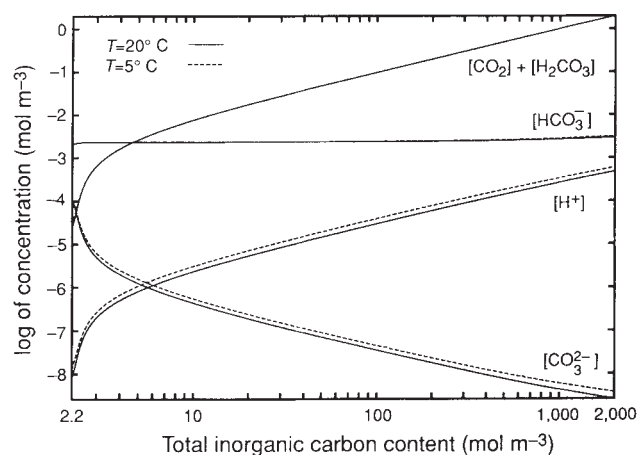


FIG. 1 Concentration of carbon constituents and H^+ for increasing C_T . The concentrations of hydrogen ion H^+ and the different forms of inorganic carbon in sea water: CO₂ (including H_2CO_3 which amounts to a few per cent of CO₂ (ref. 19)), HCO_3^- and CO_3^{2-} as functions of total inorganic carbon concentration C_T (mol m^{-3}). The $\text{pH} = -\log[\text{H}^+]$ decreases from natural level ≈ 8.2 to between 3 and 4 at maximum dissolvable C_T . $[\text{HCO}_3^-]$ stays almost constant because of constant alkalinity, which follows from assuming negligible dissolution of CaCO_3 sediments. The plot is based on apparent equilibrium constants²³ for the temperatures indicated, salinity = 35‰, $P = 30$ bar and alkalinity = 2.35 equiv. m^{-3} ; dependence on these quantities is weak for oceanic variations. The boric acid system has been included in the computations.

water depressions where accumulation of enriched water may provide a permanent threat to the environment. Furthermore, if injected CO_2 is carried towards deep-water sediments rich in CaCO_3 , dissolution of these sediments may provide efficient long-term sequestration of CO_2 as well as buffering of pH^{19} .

One possible application area could be along the west coast of Norway around 60–62° N, where water depths of 300–400 m are accessible close to the coast, and where bottom topography is sloping towards deep water in the Norwegian Sea. Furthermore, mean currents in this area are directed towards the deep ocean and the stratification is very weak below about 100 m (ref. 20) allowing easy vertical sinking of enriched water.

In other locations, strong stratification and unfavourable large-scale currents may inhibit vertical sinking and eventually lead to outgassing on a timescale of years to decades. Strong eddies or coastal upwelling may also contribute to dilute and lift CO_2 to the ocean surface in a matter of days or weeks after injection. But in many regions a starting depth of 200–400 m will be sufficient to avoid strong near-surface disturbances and ensure that the remaining ambient stratification can be penetrated by the dense CO_2 -enriched sea water.

In general, pH levels of 5 or less occur when C_T is high enough to give an appreciable density effect (Fig. 1 and equation (1)). Marine life within the area swept by the sinking bottom gravity current might therefore be severely affected. The associated temperature increase is believed to be of secondary importance. Formation of hydrates, which might create operational problems in the case of deep injection and which would sink and collect at the ocean bottom as a marine waste dump, is expected to be reduced or eliminated in the case of shallow injection (Fig. 2).

Certainly one should not recommend injection of CO_2 in the ocean to slow down the increase in greenhouse warming unless the above theoretical predictions are confirmed by experimental investigations and further analyses, and the environmental impact downstream of the injection point is deemed to be acceptable. But the chemical carrying capacity of the ocean is formidable compared with anthropogenic emissions. If shallow injection is working as indicated by these preliminary investigations, at least in some locations, it may contribute to obtaining

the required reduction in net emissions to the atmosphere in the future. □

Received 13 January; accepted 6 April 1992.

1. Marchetti, C. *Clim. Change* **1**, 59–68 (1977).
2. Baes, C. F. Jr et al. in *Interactions of Energy and Climate* (eds Bach, W., Pankrath, J. & Williams, J.) 495–519 (Reidel, Dordrecht, 1980).
3. Herzog, H., Golomb, D. & Zemba, S. *Env. Prog.* **10**, 64–74 (1991).
4. Bacastow, R. & Stegen, G. R. in *Proc. Oceans '91* Vol. 3, 1654–1657 (IEEE, Honolulu, Hawaii, 1991).
5. Stegen, G. R., Cole, K. H. & Bacastow, R. in *Proc. IEA Int. Conf. Technology Responses to Global Environmental Challenges* (IEA, Kyoto, Japan, 1991).
6. Maier-Reimer, E. in *Strategies for Future Climate Research* (ed. M. Latif) 319–339 (Max-Planck-Institut für Meteorologie, Hamburg, 1991).
7. Watanabe, H. & Izuka, K. *Metrologia* **21**, 19–26 (1985).
8. King, M. B. *Phase Equilibrium in Mixtures* (Pergamon, Oxford, 1969).
9. UNESCO *Technical Papers in Marine Science* **36** (UNESCO, Paris, 1981).
10. Smith, P. C. *Deep Sea Res.* **22**, 853–873 (1975).
11. Chu, P. C. & Gascard, J.-C. (ed.) *Deep Convection and Deep Water Formation in the Oceans* (Elsevier, Amsterdam, 1991).
12. Clift, R., Grace, J. R. & Weber, M. E. *Bubbles, Drops and Particles* (Academic, New York, 1978).
13. Tse, F. C. & Sandall, O. C. *Chem. Engng. Commun.* **3**, 147–153 (1979).
14. Calderbank, P. H., Johnson, D. S. L. & Loudon, J. *Chem. Engng. Sci.* **25**, 235–256 (1970).
15. Grace, J. R., Wairegi, T. & Brophy, J. *Can. J. Chem. Engng.* **56**, 3–8 (1978).
16. Killworth, P. D. *Deep Sea Res.* **24**, 427–448 (1977).
17. Turner, J. S. *Buoyancy Effects in Fluids* (Cambridge Univ. Press, 1973).
18. Turner, J. S. in *The Evolution of Physical Oceanography* (ed. Warren, B. A. & Wunsch, C.) 236–262 (MIT Press, Cambridge, Massachusetts, 1981).
19. Stumm, W. & Morgan, J. J. *Aquatic Chemistry* 2nd edn (Wiley, New York, 1981).
20. Otto, L. et al. *Neth. J. Sea Res.* **26**, 161–238 (1990).
21. Bishnoi, P. R. & Robinson, D. B. *Can. J. Chem. Engng.* **49**, 642–650 (1971).
22. Cheney, W. & Kincaid, D. *Numerical Mathematics and Computing* (Brooks/Cole, Monterey, California, 1985).
23. UNESCO *Technical Papers in Marine Science* **51** (UNESCO, Paris, 1987).
24. Weiss, R. F. *Mar. Chem.* **2**, 203–215 (1974).
25. Levine, I. N. *Physical Chemistry* (McGraw-Hill, New York, 1988).
26. CRC *Handbook of Chemistry and Physics* 68th Edn (ed. Weast, R. C.) (CRC Press, Boca Raton, Florida, 1987).
27. Sakai, H. et al. *Science* **248**, 1093–1096 (1990).

ACKNOWLEDGEMENTS. We acknowledge financial support from Den norske stats oljeselskap (Statoil).

Possible methane-induced polar warming in the early Eocene

L. Cirbus Sloan, James C. G. Walker, T. C. Moore Jr, David K. Rea & James C. Zachos

Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109-1063, USA

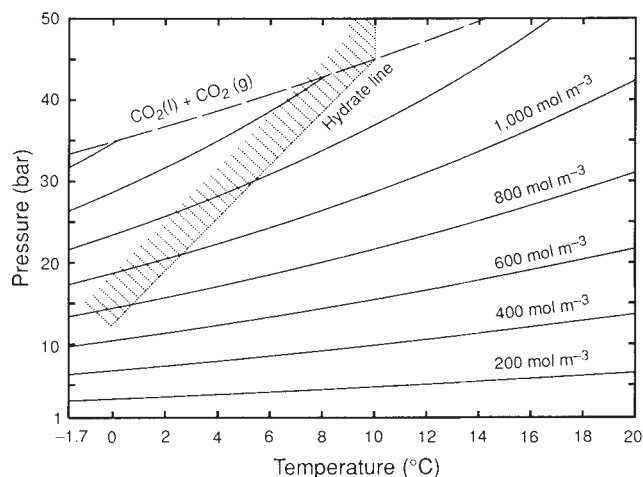


FIG. 2 Solubility of CO_2 in sea water. The number of moles of CO_2 that may be dissolved in one m^3 of sea water with salinity 35‰ (contoured), phase line between stable gaseous and liquid phase of pure CO_2 (dashed), and rough limit of conditions susceptible to hydrate formation (hatched). The pressure scale is in $\text{bar} = 10^5 \text{ Pa} = 0.9869233 \text{ atm}$; hydrostatic pressure increase in water is about 1 bar for each 10 m water depth. Maximum solubility has been calculated from a modified Henry's law, with temperature- and salinity-dependent Henry's law constant from ref. 24 and fugacity from refs 8, 25. Phase line from ref. 26. Rough limit of hydrate formation from data for pure water is discussed in ref. 8 and agrees with ref. 27.

RECONSTRUCTIONS of early Eocene climate depict a world in which the polar environments support mammals and reptiles, deciduous forests, warm oceans and rare frost conditions^{1–5}. At the same time, tropical sea surface temperatures are interpreted to have been the same as or slightly cooler than present values⁶. The question of how to warm polar regions of Earth without noticeably warming the tropics remains unresolved; increased amounts of greenhouse gases would be expected to warm all latitudes equally⁷. Oceanic heat transport has been postulated as a mechanism for heating high latitudes^{8–10}, but it is difficult to explain the dynamics that would achieve this^{7,11}. Here we consider estimates of Eocene wetland areas and suggest that the flux of methane, an important greenhouse gas, may have been substantially greater during the Eocene than at present. Elevated methane concentrations would have enhanced early Eocene global warming, and also might specifically have prevented severe winter cooling of polar regions because of the potential of atmospheric methane to promote the formation of optically thick, polar stratospheric ice clouds^{12–14}.

Methane is produced primarily by anaerobic bacteria. Major sources of methane are swamps and wetlands, rice paddies, termites and enteric fermentation by ruminants^{15–17}. For the early Eocene, anthropogenic and ruminant contributions of methane can be discounted, increasing relative contributions from wetland and forest ecosystems. There were large areas of swamps and wetlands in the late Palaeocene and early to middle Eocene, at least in part related to the long-term rise in sea level

(Fig. 1). Existing 'coaly continental clastics' of Palaeocene-Eocene age are estimated to cover an area of $2.8 \times 10^6 \text{ km}^2$ (ref. 18). These deposits are located in continental interior and coastal environments at various latitudes¹⁸. If one assumes, as a conservative estimate, that coaly deposits of a given age represent a minimum of half the total wetland environments existing at that time, the area of wetlands possible in the Palaeocene-Eocene would have been at least $5.6 \times 10^6 \text{ km}^2$, compared with $2 \times 10^6 \text{ km}^2$ for the present¹⁶. Thus during the Palaeocene-Eocene there would be the potential to triple modern methane production from wetland ecosystems alone. Additionally, there were probably more extensive areas of forest environments during the Eocene^{3,4,19} relative to present distributions, which also could have contributed to increased methane production¹⁶.

Quantitative estimates of atmospheric methane concentrations in the geological past are available only for the time recorded in ice. Thus, our hypothesis of greater methane concentrations is speculative, but is supported indirectly by a variety of evidence. We suggest that the long-term sea level rise in the late Palaeocene and early Eocene (Fig. 1) and the associated development of large areas of swamps and wetlands, along with greater areas of humid environments at this time²⁰, provided the environments for anaerobic decay of organic matter, producing increased methane flux to the atmosphere.

Sequestering of large quantities of organic, isotopically light carbon into sedimentary reservoirs should produce significant changes in global ocean ^{13}C composition (Fig. 1). The organic carbon could have been the substrate used by methanogenic bacteria to produce elevated methane fluxes. Although the relationship between late Palaeocene/early Eocene marine ^{13}C change, the carbon cycle and climate is unclear at this time, it seems that approximately concurrent with maximum ^{13}C values and the ensuing shift back towards more negative values, oxygen isotope ratios measured in benthic foraminifera shift to more negative values, indicating warmer global conditions (Fig. 2).

Methane is a powerful greenhouse gas^{15,17}. Together with enhanced carbon dioxide concentrations it could have contributed to a generally warmer troposphere and cooler lower stratosphere²¹. In addition, methane oxidation is a significant source of stratospheric water vapour^{15,17,22}. A wetter and colder lower stratosphere could have yielded polar stratospheric clouds which were optically thicker, covered a greater area and were more frequently occurring than those of today^{12-14,23}.

Optically thick polar stratospheric clouds warm the troposphere by trapping long-wave radiation from the Earth, thus preventing cooling of Earth's surface^{12-14,24,25}. Their albedo

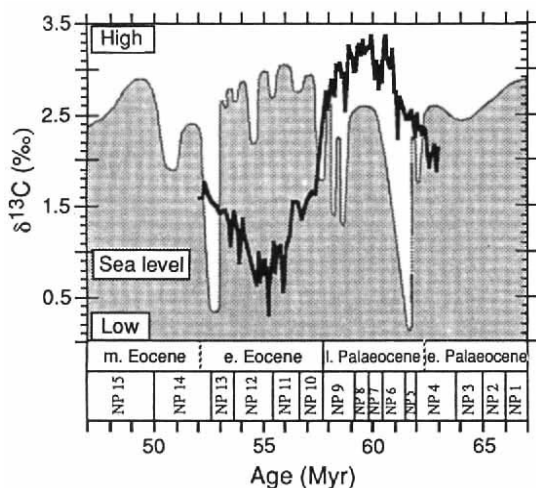


FIG. 1 Global sea level curve for the early Palaeocene³⁰ (shaded grey curve) superimposed on marine bulk carbonate ^{13}C data³¹ (solid line) for the early Palaeocene through to the middle Eocene. e, early; m, middle; l, late. The labels NP 1-15 are nanoplankton biostratigraphic zonation notation.

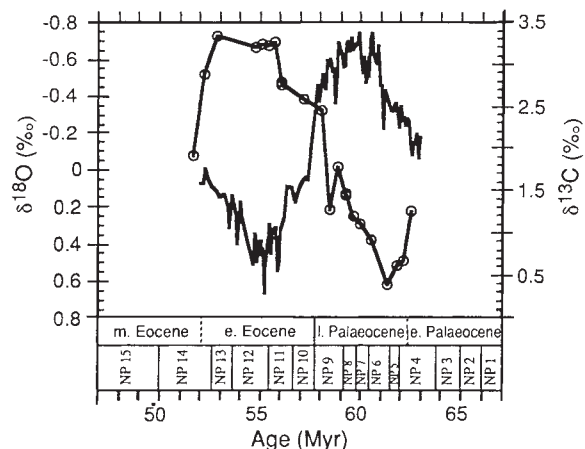


FIG. 2 Marine bulk carbonate ^{13}C data³¹ (solid line) and global oxygen curve from benthic foraminifera³² (line with circles) over the same time interval as shown in Fig. 1. e, early; m, middle; l, late.

effect on short-wave radiation is obviously small during polar winter. The net effect on the radiation balance is high-latitude warming (at present up to 0.5 K per day¹³), precisely what is needed to explain Eocene climate observations. We suggest that with a substantial methane increase and a large enough radiative effect, optically thicker, and more frequent polar stratospheric clouds with a more extensive area could potentially have kept winter polar surface temperatures above freezing, and dampened the intra-annual temperature variation at high latitudes by preventing strong winter cooling. This condition could potentially have affected Eocene deepwater structure and circulation through an influence on high-latitude surface water minimum temperatures and therefore bottom water formation³.

A methane polar stratospheric cloud system during the Eocene could have worked in conjunction with high atmospheric carbon dioxide concentration, for which there is some evidence (refs 26, 27, and K. Freeman and P. Hayes, manuscript in preparation). The response of biogenic methane sources to surface temperature changes are not well known²⁴, but warm temperatures could have enhanced methane production rates, providing a positive feedback to the system.

The photochemistry of ancient atmospheres may have been significantly different from now. The unusual climate of the Eocene may have been caused not just by enhanced greenhouse effects and modified oceanic circulation, but by enhanced methane concentrations as well. But if interpretations of early Eocene tropical sea surface temperatures⁶ are correct, then the enhanced greenhouse effect, at least in the tropics, would have had to be minor. Clearer interpretations of latitudinal surface temperature values may lead to understanding the hierarchy of the causal mechanisms for climate at this time. To describe more fully the mechanism outlined here we need more comprehensive Palaeocene-Eocene reconstructions of environments and more quantitative estimates of methane fluxes. We must also consider the interaction of atmospheric methane with other trace gases and other parts of the Palaeocene-Eocene climate system. The methane polar stratospheric cloud mechanism may also be applicable to other times when polar environments were warm, such as the Cretaceous^{28,29} and the Pliocene⁷. □

Received 3 January; accepted 6 April 1992.

- Estes, R. & Hutchison, J. H. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **30**, 325-347 (1980).
- Schweitzer, H.-J. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **30**, 297-311 (1980).
- Kennett, J. P. & Stott, L. D. *Nature* **353**, 225-229 (1991).
- Kemp, E. M. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **24**, 169-208 (1978).
- Case, J. A., in *Geology and Palaeontology of Seymour Island, Antarctic Peninsula* (eds Feldmann, R. M. & Woodburne, M. O.) 523-530 (Geol. Soc. Am., Boulder, Colorado, 1988).
- Shackleton, N. J. & Boersma, A. *Geol. Soc. Lond. J.* **138**, 153-157 (1981).
- Crowley, T. J. *Quat. Sci. Rev.* **10**, 275-282 (1991).

8. Barron, E. J. *Paleoceanography* **2**, 729–739 (1987).
9. Rind, D. & Chandler, M. J. *geophys. Res.* **96**, 7437–7461 (1991).
10. Crowley, T. J. *Paleoceanography* **6**, 387–394 (1991).
11. Sloan, L. C. & Walker, J. C. G. *Geophys. Res. Lett.* (submitted).
12. Pollack, J. B. & McKay, C. P. *J. Atmos. Sci.* **42**, 245–262 (1985).
13. Kinne, S. & Toon, O. B. *Geophys. Res. Lett.* **17**, 373–376 (1990).
14. McCormick, M. P., Trepte, C. R. & Pitts, M. C. *J. geophys. Res.* **94**, 11241–11251 (1989).
15. World Meteorological Organization/United Nations Environment Programme *Climate Change, The IPCC Scientific Assessment* (eds Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) (Cambridge Univ. Press, 1990).
16. Sheppard, J. C., Westberg, H., Hopper, J. F., Ganesan, K. & Zimmerman, P. *J. geophys. Res.* **87**, 1305–1312 (1982).
17. Cicerone, R. J. & Oremland, R. S. *Glob. Biogeochem. Cycles* **2**, 299–327 (1988).
18. Ronov, A. B., Khain, V. Y. & Balukhovskiy, A. N. *Int. Geol. Rev.* **21**, 415–446 (1979).
19. Askin, R. A. in *Geology and Paleontology of Seymour Island, Antarctic Peninsula* (eds Feldmann, R. M. & Woodburne, M. O.) 131–135 (Geol. Soc. Am., Boulder, Colorado, 1988).
20. Sloan, L. C. thesis, Pennsylvania State Univ. (1990).
21. Rasmussen, R. A. & Kahll, M. A. K. *J. geophys. Res.* **86**, 9826–9832 (1981).
22. Rinsland, C. P. et al. *J. geophys. Res.* **89**, 7259–7256 (1984).
23. McCormick, M. P. & Trepte, C. R. *J. geophys. Res.* **92**, 4297–4306 (1987).
24. Ramanathan, V. et al. *Rev. Geophys.* **25**, 1441–1482 (1987).
25. Ramanathan, V. et al. *Science* **243**, 57–59 (1989).
26. Arthur, M. A., Allard, D. & Hinga, K. R. *Geol. Soc. Am. Prog. Vol.* 23, 178 (1992).
27. Berner, R. *Am. J. Sci.* **291**, 339–376 (1991).
28. Barron, E. J. & Washington, W. M. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **40**, 103–133 (1982).
29. Wolfe, J. A. & Upchurch, G. R. Jr *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **61**, 33–77 (1987).
30. Haq, B. U., Hardenbol, J. & Vail, P. R. *Science* **235**, 1156–1167 (1987).
31. Shackleton, N. J., Hall, M. A. & Boersma, A. *Init. Rep. Deep Sea Drilling Proj.* **74**, 599–612 (US Govt Printing Office, Washington DC, 1984).
32. Miller, K. G., Janacek, T. R., Katz, M. E. & Keil, D. J. *Paleoceanography* **2**, 741–761 (1987).

TABLE 1 Characteristics of yellow-poplar leaves in 1991

CO ₂ enrichment ($\mu\text{mol mol}^{-1}$)	Specific leaf area ($\text{cm}^2 \text{g}^{-1}$)	Nitrogen (g m^{-2})	Chlorophyll (mg m^{-2})	Photo- synthesis ($\mu\text{mol m}^{-2} \text{s}^{-1}$)	Respiration ($\mu\text{mol m}^{-2} \text{s}^{-1}$)
+0	151 $\pm$ 6	1.52 $\pm$ 0.09	488 $\pm$ 54	7.4 $\pm$ 0.6	3.0 $\pm$ 0.2
+150	136 $\pm$ 5	1.30 $\pm$ 0.13	401 $\pm$ 82	10.8 $\pm$ 0.9	2.3 $\pm$ 0.2
+300	132 $\pm$ 5	1.18 $\pm$ 0.05	387 $\pm$ 32	12.3 $\pm$ 0.8	2.2 $\pm$ 0.1
P*	0.004	0.058	0.20	0.001	0.001

All data are means $\pm$ s.e. of five plants in each of two replicate chambers. Specific leaf area was determined at final harvest from total leaf area and total leaf mass. Nitrogen concentration was measured by near-infrared reflectance spectroscopy of oven-dried and ground leaves from leaf position 3–5 collected on 2 July. Chlorophyll concentrations were measured on 25 July in ethanol extracts of five 6-mm-diameter disks cut from one leaf per tree. Photosynthesis was measured with a closed gas-exchange system (LiCor 6200) on the fourth leaf from an upper branch apex. Measurements were made at midday under full sun on 1 July. Respiration was measured with the gas exchange system before sunrise on 5 July.

* Probability of a significant effect of CO₂ concentration, using sampling error in the F-test.

Productivity and compensatory responses of yellow-poplar trees in elevated CO₂

Richard J. Norby, Carla A. Gunderson,
Stan D. Wullschlegel, E. G. O'Neill
& Mary K. McCracken

Environmental Sciences Division, Oak Ridge National Laboratory,
PO Box 2008, Oak Ridge, Tennessee 37831-6034, USA

INCREASED forest growth in response to globally rising CO₂ concentrations could provide an additional sink for the excess carbon added to the atmosphere from fossil fuels^{1,2}. The response of trees to increased CO₂, however, can be expected to be modified by the interactions of other environmental resources and stresses, higher-order ecological interactions and internal feedbacks inherent in the growth of large, perennial organisms^{3,4}. To test whether short-term stimulation of tree growth by elevated CO₂ can be sustained without inputs from other environmental resources, we grew yellow-poplar (*Liriodendron tulipifera* L.) saplings for most of three growing seasons with continuous exposure to ambient or elevated concentrations of atmospheric CO₂. Despite a sustained increase in leaf-level photosynthesis and lower rates of foliar respiration in CO₂-enriched trees, whole-plant carbon storage did not increase. The absence of a significant growth response is explained by changes in carbon allocation patterns, specifically a relative decrease in leaf production and an increase in fine root production. Although these compensatory responses reduced the potential increase in carbon storage in increased CO₂ concentrations, they also favour the efficient use of resources over the longer term.

Yellow-poplar trees were grown in ambient or elevated CO₂ concentrations in open-top chambers⁵. Open-top chambers allowed the experimental plants, which before outplanting had been grown from seed under the experimental CO₂ concentrations, to be exposed in the field under ambient light, temperature, precipitation and soil conditions. The CO₂ concentration in the six chambers was maintained day and night during the growing seasons at ambient (354.5 $\mu\text{mol mol}^{-1}$), ambient +150 (502.9 $\mu\text{mol mol}^{-1}$) or ambient +300 (655.6 $\mu\text{mol mol}^{-1}$). No fertilizer or supplemental water was provided during the course of the experiment. Five yellow-poplar saplings were harvested from each chamber after 28 months (2.7 growing season).

The accumulation of above-ground biomass was similar in all treatments throughout the course of the experiment, and although stem mass was usually lowest in ambient CO₂, there were no differences in stem mass that were statistically significant (Fig. 1a). Leaf area tended to decline with increasing CO₂ concentration in 1990 and 1991, and combining the stem growth and leaf area data provided a better (less variable, more robust) indicator of the response of above-ground productivity to CO₂ enrichment (Fig. 1b). Growth efficiency, the annual production of stem mass per unit leaf area, is akin to the term net assimilation rate used with herbaceous plants and is a good indicator of general tree vigour⁶. Growth efficiency in 1990 was 41% greater

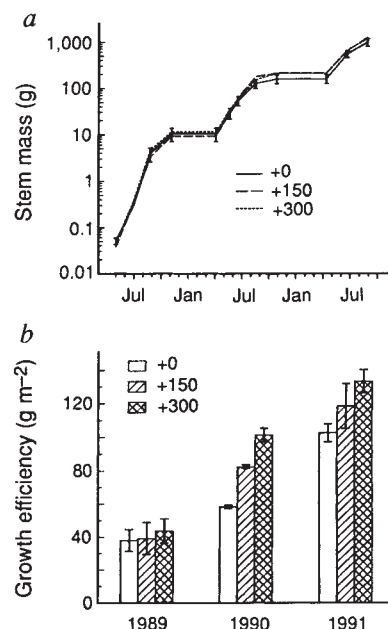


FIG. 1 Stem mass and growth efficiency of yellow-poplar saplings between May 1989 and August 1991 in ambient (+0), ambient +150, and ambient +300 $\mu\text{mol CO}_2 \text{mol}^{-1}$. All data are the means $\pm$ s.e. of five trees per chamber in two replicate chambers. a, Stem dry mass was estimated from periodic measurements of bole diameter (D) 10 cm from the ground and total height (H) of the bole, using a log-log regression between stem mass and D^2H . b, Growth efficiency was calculated as annual stem mass increment divided by leaf area at the end of the year. The effect of CO₂ on growth efficiency was significant at $P=0.041$ in 1990, and $P=0.002$ in 1991.

TABLE 2 Fine root mass and CO₂ efflux from the ground within open-top chambers

CO ₂ enrichment ($\mu\text{mol mol}^{-1}$)	Root mass density (g m^{-2})		CO ₂ efflux ($\mu\text{mol m}^{-2} \text{s}^{-1}$)
	Live	Dead	
+0	85.0 $\pm$ 22.3	51.6 $\pm$ 15.5	6.60 $\pm$ 0.53
+150	157.4 $\pm$ 31.3	61.3 $\pm$ 11.0	8.21 $\pm$ 0.72
+300	186.2 $\pm$ 53.6	51.0 $\pm$ 10.2	8.03 $\pm$ 0.55
P*	0.032	NS	0.150

Fine roots (<7 mm diameter) were quantified in five cores (10 cm diameter $\times$ 15 cm deep) per chamber collected 4 days after tree boles were removed. Data are the means $\pm$ s.e. of two replicate chambers. CO₂ efflux from 560 cm² of ground surface was measured on 11 August 1991 with a modified LiCor 6200 gas exchange system (P. J. Hanson, personal communication). Data are the means $\pm$ s.e. of three to four measurements per chamber.

* Probability of a significant effect of CO₂ concentration, using sampling error in the *F*-test. NS, not significant ($P > 0.20$).

in +150 and 74% greater in +300 relative to +0 chambers. From April to August 1991, growth efficiency was enhanced by 16% in +150 and 33% in +300, consistent with the expected response of plants to CO₂ enrichment^{3,7}.

Increased growth efficiency of yellow-poplar trees in CO₂-enriched atmospheres can be attributed to the sustained increase in photosynthesis we have observed in these trees. Light-saturated photosynthesis of upper canopy leaves on 1 July 1991 was 46% higher in +150 and 67% higher in +300 compared with that of leaves in +0 (Table 1); these values are representative of the response of photosynthesis throughout the experiment. The relative responses of photosynthesis to CO₂ enrichment were similar in older and shaded leaves from different parts of the canopy, and there was no effect of CO₂ on canopy longevity (bud break or autumn senescence). Stimulation of photosynthesis was sustained even though nitrogen concentration and chlorophyll concentration in leaves decreased with increasing CO₂ (Table 1). Coupling the increased photosynthesis with lower rates of foliar respiration (Table 1) implies a substantial increase in carbon uptake in elevated CO₂ concentrations⁵.

The juxtaposition of increased carbon uptake and increased growth efficiency with no significant effect on stem mass is only partly explained by reduced leaf area (10% lower in +300); other changes in the patterns of carbon allocation to biomass and carbon losses from the plant are implied. When the trees were harvested in August 1991, however, we found no significant effects of CO₂ on the mass of the bole, primary or secondary branches, nor on the proportionate distribution of stem mass between the lower, middle and upper parts of the trees (Fig. 2). There were no indications that canopy structure of these yellow-poplar saplings was altered: tree height, diameter, the number and length of primary branches, and the ratio of branch mass to mass of the bole were similar in all treatments. The tap root was the only harvested component of the trees that significantly increased in mass in response to CO₂ enrichment, increasing 12% and 37% in +150 and +300 relative to +0 ($P = 0.072$). But there was no effect of CO₂ on total root mass or root-to-shoot ratio.

Leaf area ratio (leaf area divided by whole-plant mass) declined significantly and linearly with increasing CO₂. Leaf area ratio was 12% lower in +150 and 21% lower in +300 compared with +0. Leaf area ratio represents the balance between the assimilation capacity (leaf area) and respiratory demand (plant mass); relative growth rate will tend to decline as leaf area ratio declines unless there is a compensating increase in leaf efficiency (for example, photosynthetic stimulation). The maintenance or enhancement of growth in high CO₂ without a concomitant increase in leaf area represents a mechanism for increasing whole-plant water-use efficiency⁴, a response that could improve environmental fitness over the long term.

The indeterminate nature of leaf production in yellow-poplar provides the opportunity for continuous adjustments in carbon allocation patterns to optimize the use of limiting environmental resources. In these trees, as in many forest trees, nitrogen was a limiting resource, and the concentrations of nitrogen and chlorophyll in leaves declined with CO₂ enrichment (Table 1). A common response to nitrogen limitations is a shift in allocation from leaf production to root production⁶. The mass of live fine roots (<7 mm diameter) increased significantly with CO₂ enrichment (Table 2). Yellow-poplar fine roots turn over rapidly, accounting for a significant loss of fixed carbon from the system⁸. CO₂ efflux from the soil (including lateral roots) in the chambers was indeed higher in elevated CO₂ (Table 2). Hence, the compensatory response in carbon allocation between leaf and fine root production, which may be associated with an enhancement in resource utilization, may effect short-term reductions in carbon storage by the system.

The results of this experiment, the longest so far in which any forest tree species has been exposed to elevated CO₂, show that the short-term responses of yellow-poplar seedlings to increased CO₂ (ref. 4) can be sustained over several growing seasons under field conditions: photosynthesis and growth efficiency remained significantly higher in CO₂-enriched saplings during the third year of enrichment. Nevertheless, there was no significant enhancement of tree dry mass (carbon storage) in this system. The leaf-level gas-exchange responses, and the more integrative response of growth efficiency, did not effect an increase in whole-plant carbon storage because of the changes in carbon allocation patterns between above-ground (leaf production) and below-ground (fine-root production) processes. This shift in allocation represents a trade-off between potential carbon acquisition versus water and nutrient acquisition, as well as a shift from carbon storage in perennial tissue versus production of the more ephemeral fine roots.

These results contrast with the more spectacular response of irrigated and fertilized sour orange trees (*Citrus aurantium*) to

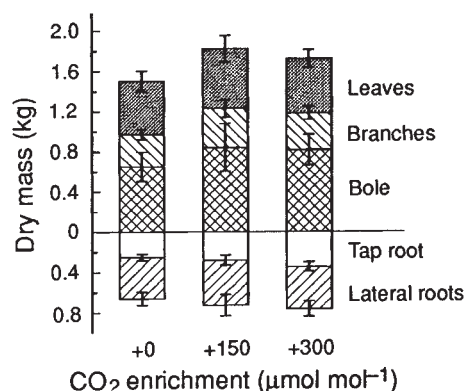


FIG. 2 Mass of components of yellow-poplar saplings when harvested in August 1991. Data are the means $\pm$ s.e. of five plants in each of two replicate chambers per CO₂ concentration. The only component for which there was a statistically significant effect of CO₂ was tap root ($P = 0.07$). Harvesting involved separating the stems and associated leaves into the bole, primary and secondary branches. Branches arising from the lowest 1 m of the trunk were cut off and processed first, followed by branches from the second metre of trunk, and then the topmost length of bole and its branches. The remaining bole was then cut off at the ground and separated into two 1-m sections. Leaves and stems were oven-dried (70 °C) and weighed. The tap-root system was excavated by digging a cylinder around the stump of ~ 25 cm in diameter, then severing all lateral roots to a depth of 25 to 30 cm. The tap root was then pried out of the ground with little breakage. The tap root and all broken roots in the cylinder of soil were washed free of adhering soil. The diameter of each lateral root was measured at the point where it left the excavated cylinder. Eight lateral root systems, ranging in initial diameter from 4 to 26 mm, were excavated, washed and oven-dried to establish a regression between dry mass and diameter, from which the mass of lateral roots of each tree was calculated.

elevated CO₂, from which conclusions about the response of the global forest have been drawn⁹. Such extrapolations from a single species growing under conditions uncharacteristic of a forest are neither justified nor particularly relevant to global carbon cycling. Our yellow-poplar responses provide evidence that the dramatic increases in tree growth reported for sour orange trees will not occur in all systems, but it should not be concluded from this that forests cannot provide a larger sink for increasing concentrations of atmospheric CO₂. Increased growth efficiency in elevated CO₂ implies that the trees may be better able to withstand environmental stresses because less leaf area is required to sustain the same growth rate. The compensatory responses in leaf and fine root production should also favour efficient use of resources over the long term. The potential for a forest ecosystem to provide a sink for atmospheric carbon may depend more on reactions to environmental stresses and resource availability than on the direct responses of growth rate to CO₂ concentration. Recent simulations with a forest stand succession model show that when the indirect effects of elevated CO₂ on carbon, nitrogen, and water dynamics are considered, considerable potential exists for increasing carbon storage in forest ecosystems (W. M. Post, personal communication). These results with yellow-poplar trees, then, emphasize the importance of resource interactions and feedbacks in the analysis of forest responses to rising CO₂ concentrations. □

Received 10 January; accepted 2 April 1992.

1. Tans, P. P., Fung, I. Y. & Takahashi, T. *Science* **247**, 1431–1438 (1990).
2. Post, W. M. *et al. Am. Sci.* **78**, 310–326 (1990).
3. Bazzaz, F. A. *A. Rev. Ecol. Syst.* **21**, 167–196 (1990).
4. Norby, R. J. & O'Neill, E. G. *New Phytol.* **117**, 515–528 (1991).
5. Wullschlegel, S. D., Norby, R. J. & Gunderson, C. A. *New Phytol.* (in the press).
6. Waring, R. H. & Schlesinger, W. H. *Forest Ecosystems: Concepts and Management* (Academic, Orlando, Florida, 1985).
7. Eamus, D. & Jarvis, P. G. *Adv. ecol. Res.* **19**, 1–55 (1989).
8. Harris, W. F., Sollins, P., Edwards, N. T., Dinger, B. E. & Shugart, H. H. in *Productivity of World Ecosystems* (Symp. Proc. International Biological Program, Seattle, 31 August–1 September 1972) 116–122 (Natl. Acad. Sci., Washington DC, 1975).
9. Idso, S. B. *Bull. Am. meteorol. Soc.* **72**, 962–965 (1991).

ACKNOWLEDGEMENTS. We thank J. Arnold, M. Cockrill, F. Figarella, T. Grizzard, P. Hanson, G. Hoffman, J. Jackson, A. King, J. Rasnake, J. Riggs, B. Selvidge and R. Stewart for their participation in this project. This work was sponsored by the Carbon Dioxide Research Program, Office of Health and Environmental Research, US Department of Energy.

Earliest known simian primate found in Algeria

Marc Godinot* & Mohamed Mahboubi†

* Institut des Sciences de l'Evolution (CNRS, URA 327), Case Courrier 64, Université Montpellier-2, Place Eugène-Bataillon, 34095 Montpellier Cedex 05, France

† Institut des Sciences de la Terre, Université d'Oran, BP 16, Oran-Es Sénia, Algeria

THE record of early fossil Simiiformes (Anthropoidea¹) from the Late Eocene and Early Oligocene of Africa and the Arabian Peninsula has increased dramatically in recent years^{2–6}. We report here the discovery of a new, diminutive and much older (Early or Middle Eocene) simian from an Algerian locality, Glib Zegdou. This species is smaller than any other living or fossil African simiiform. Derived similarities shared with *Aegyptopithecus* suggest that the new genus is more closely related to propithecines than to oligopithecines, implying that these two subfamilies differentiated during the Early Eocene. The new discovery confirms predictions about the great antiquity of Simiiformes^{7–9} and emphasizes a long and endemic African history for higher primates.

Isolated mammalian teeth were recovered from the upper fossiliferous level of the Glib Zegdou section, which had previously yielded charophytes¹⁰. The Glib Formation is situated

in the northern part of the Hammada du Dra, Algerian Sahara. The mammals include hyracoids, rodents, primates and other new taxa, probably condylarthrans. Five primate teeth have been recovered, of which two well preserved molars and one premolar represent the new simian, described below. The two others are less well preserved and represent a second, slightly larger and related species.

For the few mammals from other Glib Zegdou loci and from the nearby correlative Gour Lazib, a Middle Eocene age has been proposed^{11,12}. Early and Middle Eocene mammals are poorly known in Africa, however. Charophytes (algal cysts) provisionally remain a better basis for a continental biostratigraphy. They have permitted estimation of an Early Eocene age for Glib Zegdou¹⁰. Work in progress (M. Feist and F. Mebrouk, personal communication) confirms a late Early Eocene age (46–50 Myr), which we favour here, but more evidence is still needed to exclude definitely a Middle Eocene (40–46 Myr) age.

Algeripithecus minutus gen. et sp. nov.

Holotype. Left upper molar M², GZC 1 (Oran University collection; Fig. 1a).

Referred material. Right lower molar M₃ (GZC 2; Fig. 1b) and right lower premolar P₃ (GZC 3).

Horizon and locality. A 50-cm sandstone bed at the top of the lower third of the Glib Formation at Glib Zegdou, Algeria.

Diagnosis. Small size (M²: length, 1.66 mm; width, 2.63 mm. M₃: length, 2.08 mm; width, 1.37 mm). M² transversely elongated, bearing a large cuspidate hypocone, a lingual cingulum and an incipient pericone. Labial slope of paracone and metacone, and lingual slope of protocone and hypocone all very low (lower than in any known adapiform or omomyiform). M₃ broad, moderately high-crowned and bunodont. Trigonid open anteriorly, with a thin ventro-lingually oriented paracristid, and a shorter and much thicker premetacristid. Protoconid and metaconid close and connected by a high protocristid. M₃ buccal cingulid continuous from the hypoconulid to the anterior rim.

Although this material is fragmentary, its remarkable characteristics lend it far-reaching significance. Foremost is the very small size of the teeth, which are smaller than those of *Catopithecus* and *Proteopithecus*, recently described from the Fayum of Egypt⁵, and slightly smaller than those of *Qatrania wingi* and *Biretia piveteaui*, the smallest previously known African simians^{4,13,14}. We suggest a broad body weight estimate of 150 to 300 g, based on tooth size-body size regressions for generalized living primates¹⁵ (mean weight from M₃ is 246 g, and mean from M² is 285 g). *Algeripithecus* hence was in the size range of New World callitrichids and small prosimians.

Among fossil Euprimates, species of similar size are known

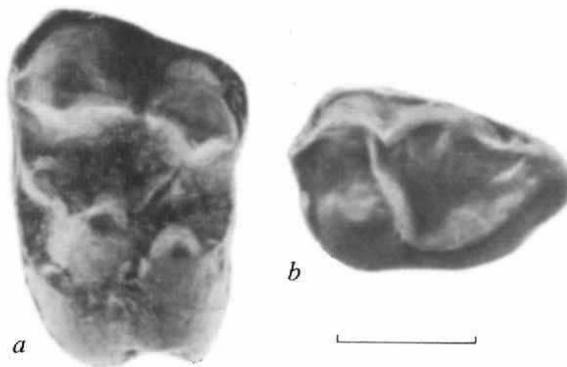


FIG. 1 Scanning electron micrographs of the two molars of *Algeripithecus minutus* gen. et sp. nov. from Glib Zegdou, in occlusal views. a, The type-specimen, left M² GZC 1; b, the right lower M₃ GZC 2; scale bar, 1 mm. (Micrographs by A. Rossi).

among Eocene omomyids, adapiforms and parapithecids. Comparisons reveal a superficial similarity with the already larger Middle Eocene European adapiform *Periconodon*, which also possesses transversely expanded upper molars, with moderate to large hypocone and pericone¹⁶. However, *Algeripithecus* differs from *Periconodon* in its lower relief, more buccally placed hypocone on M^2 (Fig. 2), and its less elongated and more triangular M_3 , which sets it apart from known adapiforms. Moreover, *Algeripithecus* shows peculiar characters that are unknown in any adapiform or omomyid, including labial expansion at the base of both paracone and metacone and the very low lingual slope of the protocone and hypocone. These characters result in a distinctive upper molar pattern, which is rare among primates, but present in the much larger simian *Aegyptopithecus*, from the Fayum fauna (Fig. 2). The hypocone of *Algeripithecus*, large, bulbous with a small anterior crest, is different from that of the oligopithecines, but very similar to that of *Aegyptopithecus*¹⁷. The uniqueness of this combination of derived characters makes it highly probable that they are homologous in these two genera. An ancestor-descendant relationship seems excluded by the discontinuous M_3 paralophid, and possibly by precociously extreme bunodonty in *Algeripithecus*. However, a sister-group relationship between them is proposed (Fig. 3).

Other interpretations are much less likely. Possible changes in interpretation between homologous and convergent characters push oligopithecines further away in simian relationships without altering the sister-group relationship between *Algeripithecus* and propliopithecines (Fig. 3b). Convergent acquisition by *Algeripithecus* and *Aegyptopithecus* of their similarities, which are so unique in fossil primates, is too unpar-

simonious. Considering *Algeripithecus* to approach the ancestral morphology of the group would imply highly unlikely reversals in oligopithecines (trigon basin elongation in *Proteopithecus*, re-increase of shearing crests), and unwarranted changes (from large cuspidate to crestiform hypocone). We reject such hypotheses. The phylogeny advocated here (Fig. 3a) implies that propliopithecines and oligopithecines split during the Early Eocene. Such an ancient history had already been suggested by the major morphological differences between them⁵. It also implies that the earlier dichotomies in simian phylogeny also took place at a very early date: parapithecids and platyrrhines must have differentiated during the Early Eocene, or before.

The small size and advanced bunodonty of *Algeripithecus* has consequences for simiiform origins. In living primates, small size is linked with increased insectivory and high-crested teeth¹⁸, a trend which is also known in the smallest adapiform *Anchomomys* and in the tiny omomyiform *Pseudoloris*^{16,19}. This general trend makes it very unlikely that *Algeripithecus* could be derived in a short time interval from any known adapiform, including the oldest and most primitive *Donrussellia provincialis*²⁰. This last species, from the earliest Eocene, is already slightly larger than *A. minutus* (Fig. 2). *Algeripithecus*, being

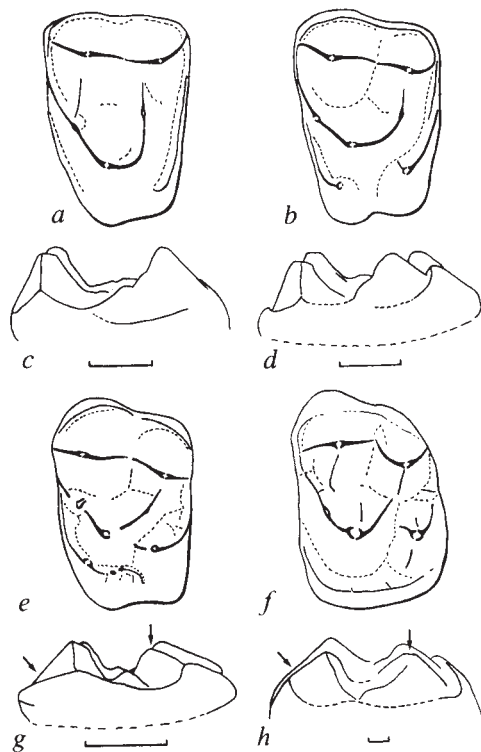


FIG. 2 Comparative schematic drawings of the M^2 of four Palaeogene primates: the adapiforms *Donrussellia provincialis* (a and c, Paris MNHN RI 171) and *Periconodon helveticus* (b and d, Basel Ef 366), the new simian *Algeripithecus minutus* (e and g, Oran GZC 1) and the propliopithecine *Aegyptopithecus zeuxis* (f and h, Cairo Geological Museum 40237). All teeth drawn to the same length; scale bars, 1 mm; a, b, e and f, occlusal views; c, d, g and h, anterior views. Arrows point towards the low protocone and the buccal expansion of the paracone in the two simians.

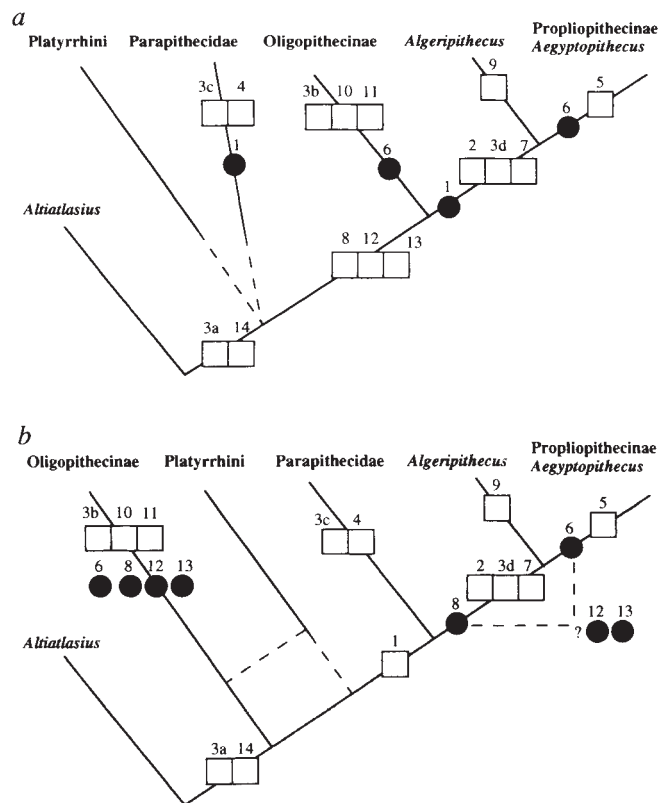


FIG. 3 Two cladograms of early simiiform relationships, based on the analysis of 14 dental traits. Derived characters are numbered and indicated as shared derived (squares) or convergent (filled circles). Cladogram a requires fewer convergences than b, and is therefore favoured by us. 1, Marked bunodonty, decrease in shearing crests; 2, labial expansion of paracone and metacone on M^2 ; 3a, small and crestiform hypocone; 3b, large and crestiform hypocone; 3c, large cuspidate hypocone, broadly linked with the lingual cinogulum; 3d, large cuspidate hypocone, with a small anterior crest towards the protocone wall and having lost a major link with the lingual cinogulum; 4, enlarged conules; 5, much larger size; 6, reduction of paraconule; 7, M_3 buccal cingulid continuous between hypoconid and hypoconulid; 8, reduction of metaconule; 9, discontinuous paralophid on M_3 ; 10, reduction of lingual lobe of P_3 ; 11, twinned entoconid and hypoconulid; 12, loss of P_2 ; 13, similar broadening of P_3 and molarization of P_4 ; 14, some paraconid reduction on lower molars. (Characters 12 and 13, being unknown in *Algeripithecus*, have two alternative locations on cladogram b.)

much more bunodont than similarly sized omomyids and adapiforms, suggests that early simians had different dietary habits, probably including a larger proportion of hard food. This adaptation must date back to the Early Eocene, or Palaeocene, and is probably linked with the early acquisition of simiiform skull characters. The Early (or Middle) Eocene age of *Algeripithecus* definitely rules out the hypothesis that Middle and Late Eocene European adapiforms are the stem group for simiiforms^{2,5,21-23}.

The existence of a distinct and ancient euprimate group in Africa is documented by the Palaeocene *Altiasius* from Morocco²⁴. This genus is more bunodont than other omomyids²⁴, and on M² bears a continuous lingual cingulum with incipient hypocone and pericone. These two derived characters shared with later simians suggest a very likely relationship between them. This would imply a Palaeocene (53–65 Myr) split between strepsirhines and simiiforms. Regardless of the relationships of *Altiasius*, *Algeripithecus* demonstrates the great antiquity of Simiiformes, and of their basal dichotomies, in Africa. This in turn lengthens the time interval for a chance dispersal of platyrrhines across the Atlantic ocean, increasing the likelihood of simian monophyly^{1,7,8,25,26}. □

Received 21 November 1991; accepted 10 April 1992.

- Hoffstetter, R. *Ann. Paleont.* **68**, 241–290 (1982).
- Rasmussen, D. T. & Simons, E. L. *Folia Primatol.* **51**, 182–208 (1988).
- Thomas, H., Roger, J., Sen, S. & Al-Sulaimani, Z. *C. r. hebdom. Acad. Sci. Ser. II* **306**, 823–829 (1988).
- Bonis, L. de, Jaeger, J.-J., Coiffait, B. & Coiffait, P.-E. *C. r. hebdom. Acad. Sci. Ser. II* **306**, 929–934 (1988).
- Simons, E. L. *Proc. natn. Acad. Sci. U.S.A.* **66**, 9956–9960 (1989).
- Simons, E. L. *Science* **247**, 1567–1569 (1990).
- Hoffstetter, R. *Bull. Mem. Soc. Anthropol. Ser. XIII*, **4**, 327–346 (1977).
- Martin, R. D. *Primate Origins and Evolution* (Chapman and Hall, London, 1990).
- Beard, K. C., Krishtalka, L. & Stucky, R. K. *Nature* **349**, 64–67 (1991).
- Gevin, P., Feist, M. & Mongereau, N. *Bull. Soc. Hist. nat. Afrique du Nord* **65**, 371–376 (1974).
- Sudre, J. C. *r. hebdom. Acad. Sci. Ser. D* **280**, 1539–1542 (1975).
- Sudre, J. *Palaeovertebrata* **9**, 83–115 (1979).
- Simons, E. L. & Kay, R. F. *Nature* **304**, 624–626 (1983).
- Simons, E. L. & Kay, R. F. *Am. J. Primatol.* **15**, 337–347 (1988).
- Gingerich, P. D., Smith, B. H. & Rosenberg, K. *Am. J. Phys. Anthropol.* **58**, 81–100 (1982).
- Godinot, M. *Cour. Forsch.-Inst. Senckenberg* **107**, 383–407 (1988).
- Fleagle, J. G. & Kay, R. F. *J. hum. Evol.* **16**, 483–532 (1987).
- Kay, R. F. & Covert, H. H. in *Food Acquisition and Processing in Primates* (eds Chivers, D. J., Wood, B. A. & Bilsborough, A.) 467–508 (Plenum, New York, 1984).
- Godinot, M. *Palaeontographica* **205**, 113–127 (1988).
- Godinot, M. *Palaeovertebrata* **10**, 43–126 (1981).
- Gingerich, P. D. *Geobios Mem. Spec.* **1**, 165–182 (1977).
- Gingerich, P. D. in *Evolutionary Biology of the New World Monkeys and Continental Drift* (eds Ciochon, R. L. & Chiarelli, A. B.) 123–138 (Plenum, New York, 1980).
- Rasmussen, D. T. *Int. J. Primatol.* **11**, 439–468 (1990).
- Sigé, B., Jaeger, J.-J., Sudre, J. & Vianey-Liaud, M. *Palaeontographica* **214**, 31–56 (1990).
- Szalay, F. S. & Delson, E. *Evolutionary History of the Primates* (Academic, New York, 1979).
- Fleagle, J. G. *Primate Adaptation and Evolution* (Academic, New York, 1988).

Unusual *HLA-B* alleles in two tribes of Brazilian Indians

Mônica P. Belich[†], J. Alejandro Madrigal^{*}, William H. Hildebrand^{*}, Jacqueline Zemmour^{*}, Robert C. Williams[‡], Roberto Luz[†], Maria Luiza Petzl-Erler[†] & Peter Parham^{*}

^{*} Departments of Cell Biology, Microbiology and Immunology, Stanford University, Stanford, California 94305, USA

[†] Department of Genetics, Federal University of Paraná, 81531 Curitiba, Paraná, Brazil

[‡] Histocompatibility Laboratory, Blood Systems Inc., 6220 East Oak Street, Scottsdale, Arizona 85257, USA

THE Kaingang and Guarani are culturally and linguistically distinct tribes of southern Brazil^{1,2}. Like all Amerindian groups^{3,4} they show limited *HLA* polymorphism, which probably reflects the small founder populations that colonized America by overland migration from Asia 11,000–40,000 years ago^{5,6}. We find the nucleotide sequences of *HLA-B* alleles from the Kaingang and Guarani to be distinct from those characterized in caucasian, oriental and other populations⁷. By comparison, the *HLA-A* and *C* alleles are familiar. These results and those reported in the accompanying paper⁸ on the Waorani of Ecuador reveal that a marked evolution of *HLA-B* has occurred since humans first entered South America. New alleles have been formed through recombination between pre-existing alleles, not by point mutation, giving rise to distinctive diversification of *HLA-B* in different South American Indian tribes.

Serological *HLA-A,B,C* typing showed the Kaingang and Guarani have typically Amerindian patterns³ of *HLA-A,B,C* antigens (Table 1). A few antigens, three to five, dominate each locus and certain antigens, for example the A10 family, are characteristically absent. Differences between the Kaingang and Guarani were also apparent. Few *HLA-A,B* alleles typed as *HLA* blank, although certain patterns of serological reactivity, notably B*53' in the Guarani, were unusual. For *HLA-C* the frequency of blanks was higher, which is true for most populations⁹.

Epstein Barr virus-transformed cell lines were established from representative Kaingang and Guarani donors. Isoelectric focusing demonstrated the presence of *HLA-A2,B35,B40* and *B62* molecules distinct from those in the caucasian population. Nucleotide sequencing of *HLA-A,B* and *C* alleles from cell

TABLE 1 Serological allele frequencies for *HLA-A,B* and *C* in Guarani and Kaingang Amerindians

Guarani (%)		Kaingang (%)	
A2	52.6	A31	64.8
A28	32.7	A2	19.4
A31	11.2	A24	6.2
A30	1.5	A29	5.2
A24	1.0	A11	2.2
A11	0.5	A26	1.3
A29	0.5	A3	0.4
A blank	0	A blank	0.5
B40	40.8	B35	34.4
B15(62)	28.1	B51	25.3
B35	20.9	B39	20.5
B*53'	6.6	B40	8.5
B8	1.5	B13	4.0
B39	1.5	B15(62)	3.1
B51	0.5	B14	1.3
B blank	0.1	B37	0.4
		B blank	2.5
Cw3	83.3	Cw4	37.3
Cw4	6.6	Cw7	20.1
Cw7	2.6	Cw1	11.3
Cw5	0.5	Cw3	10.9
C blank	6.5	Cw6	4.4
		Cw8	1.3
		Cw2	0.9
		C blank	13.8

Peripheral blood lymphocytes from 98 Guarani and 113 Kaingang from the Rio das Cobras reservation in the state of Paraná were typed with a panel of 146 anti-*HLA-A,B,C* alloantisera. These sera were obtained from laboratories in several countries and from a screening programme in Curitiba. They were selected to emphasize the antigenic specificities most often detected in Amerindians and were screened with a reference panel of non-Amerindian cells to check their specificity and to allow comparison of the patterns of reactivity with the Kaingang and Guarani cells. B15 has been subdivided into a number of component antigens, of which B62 is the most common.

lines expressing these electrophoretic 'variants' revealed additional 'variant' alleles not detected by isoelectric focusing, particularly for *HLA-B* (Table 2).

Seven of the *HLA-B* alleles analysed all gave previously undiscovered sequences: two alleles were common to the tribes, two were found only in the Kaingang and three only in the Guarani (Fig. 1). Six of the alleles differ from known *HLA-B*

alleles^{7,10} by small clusters of nucleotide substitutions, which localize to the first third of exon 3 (Fig. 2) and mostly produce substitutions in the peptide-binding residues of the β strands of the α_2 domain (Fig. 1). In each case the cluster of differences can be found in other *HLA-B* alleles and in most instances in another allele present in the tribe. Such patterns indicate that the six *HLA-B* alleles were formed by single microrecombination events between pre-existing *HLA-B* alleles.

The seventh *HLA-B* allele encodes a chimaera in which a

part of the α_1 domain of B*3901 has been introduced into B*4001. Additional differences in the leader sequence and at positions 97 and 245 (Fig. 3) suggest the evolutionary history of this *HLA-B* allele is more complex than for the other six. Subsequently we found it to be identical to that encoding the B48 antigen (B*4801), which was isolated from native Americans of the Pima tribe of Arizona. *HLA-B48* is rare throughout most of the world but has high frequencies in Eskimos and other North American Indians, for example 33% in the Pima¹¹. Of

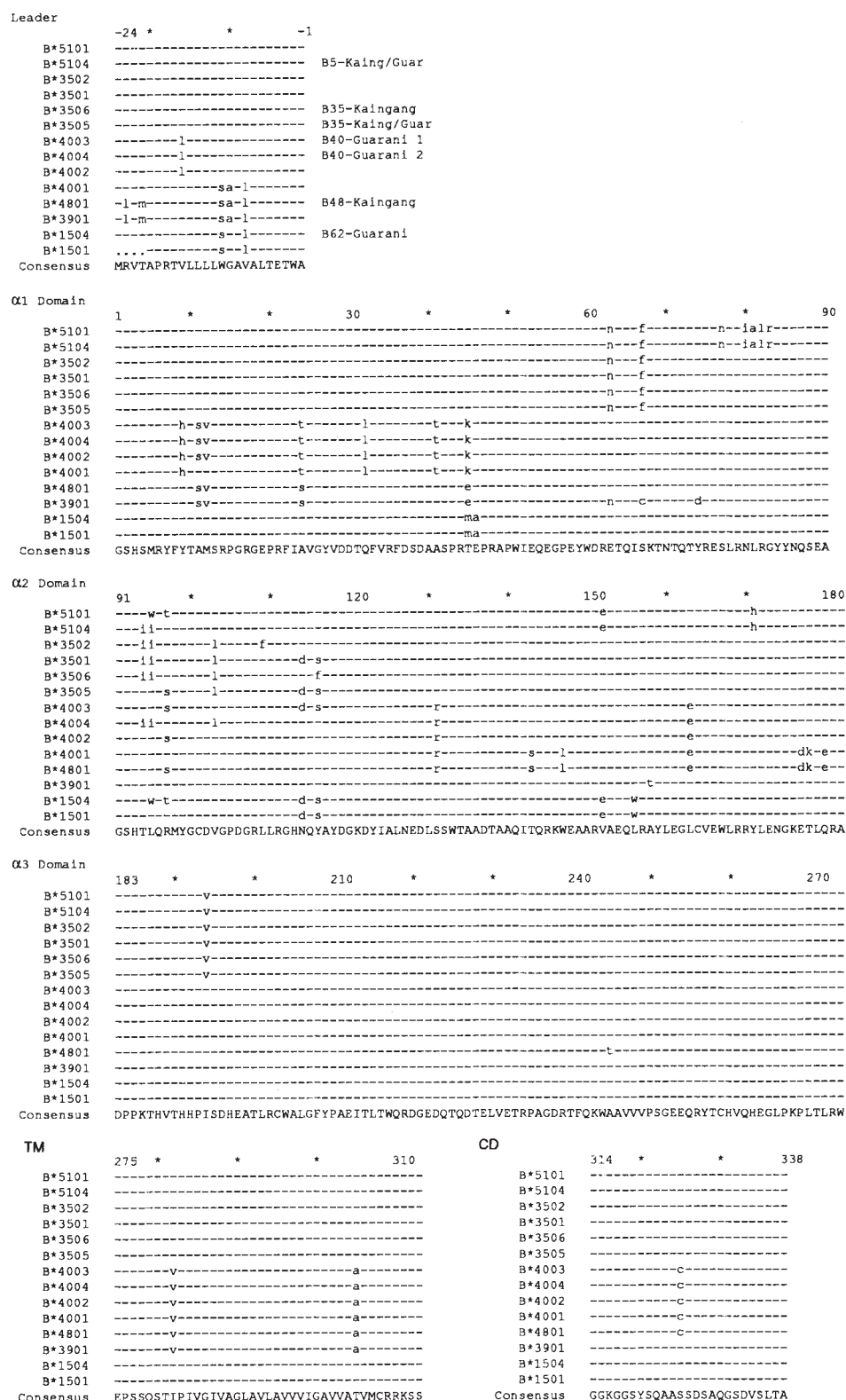


FIG. 1 Predicted protein sequences of the heavy chains encoded by Kaingang and Guarani *HLA-B* alleles are compared to related *HLA-B* alleles from other populations. Official WHO nomenclature for the alleles is given on the left¹¹. TM, transmembrane domain; CD, cytoplasmic domain. *HLA-A,B* alleles were isolated and characterized as described^{23,24}.

particular interest is that B*4801 has a unique threonine residue at position 245, which from previous analysis of A2 and A68 might be expected to reduce affinity of the B48 protein for the CD8 co-receptor of T cells¹².

Unlike *HLA-B*, four of five Kaingang and Guarani *HLA-A* alleles analysed were not unusual (Table 2). *HLA-A2* and *A24* alleles identical to the common caucasian forms were isolated from the Guarani, and *A31* from the Kaingang was only distinguished from the caucasian form by a single silent substitution. A second subtype of A2 from the Guarani is identical to the A*0211 oriental subtype¹³. The only 'novel' *HLA-A* molecule from the two Brazilian tribes was an A2 subtype from the Kaingang (A*0212). It differs from the common A*0201 subtype by substitution of glutamine for leucine at position 156 and could have been formed by recombination with *A31*. Of two Kaingang *HLA-C* alleles, one is identical to that encoding caucasian Cw4; the other is a new subtype of Cw8, differing by one amino acid (arginine for glycine at position 175) from the Cw*0801 allele found in the Pima¹⁴. This replacement is due to a unique nucleotide substitution probably resulting from a point mutation.

In studying the Waorani Indians of Ecuador, Watkins *et al.*⁸ find *HLA-B* to be similarly divergent, whereas that was not the case for the North American Zuni. Five of the seven Kaingang and Guarani *B* alleles are not shared with the Waorani (B*1504 and B*4801 are shared) and conversely there are Waorani *B* alleles not found in the Brazilian tribes. For example, different subtypes of B35 are found in the Brazilian and Ecuadorian tribes, whereas the Zuni of North America have the same B*3501 subtype as the Japanese. This pattern of *HLA-B* diversification between tribes (private polymorphisms¹⁵) is reminiscent of class I H-2 antigen differences between populations of wild mice¹⁶.

Their absence in other populations, including North American Indians, suggests the unusual *HLA-B* alleles of the Kaingang, Guarani and Waorani are of South American origin and that *HLA-B* has undergone considerable diversification in South America in the last 11,000–40,000 years. Consistent with a recent evolution for these alleles are their simple structural relationships, which emphasize interallelic recombination or conversion as the principal factor in *HLA-B* diversification¹⁷. Although not

TABLE 2 Alleles sequenced From Kaingang and Guarani

Allele	WHO nomenclature	Cell line	Serological HLA type†
A2 Kaingang	A*0212	KRC 033	A2,31: B39,35: Cw7,w4
A2 Guarani 1‡	A*0201	GRC 138	A2,2: B62,40: Cw3,w3
A2 Guarani 2‡	A*0211	GRC 138	
A24 Kaingang‡	A*2402	KRC 032	A24,31: B35,35: Cw4,w4
A31 Kaingang‡	A*31012	KRC 033	
B35 Kaingang/Guarani	B*3505	KRC 032	
		GRC 212	A2,28: B35,40: Cw4,w3
B35 Kaingang	B*3506	KRC 103	A31,31: B39,40: Cw7, —
B40 Guarani 1	B*4003	GRC 138	
B40 Guarani 2	B*4004	GRC 212	
B48 Kaingang‡§	B*4801	KRC 103	
B5 Kaingang/Guarani	B*5104	KRC 005	A2,A2: B51,51: Cw1,w1
		GRC 187	A28,31: B35,62: Cw3,w3
		GRC 150	A31,A2: B5,62: Cw3, —
B62 Guarani‡	B*1504	GRC 138	
Cw4 Kaingang‡	Cw*0401	KRC 103	
Cw8 Kaingang	Cw*0803	KRC 103	

† Owing to the unusual nature of the *HLA-B* alleles, the serological types and the overall structural relationships of the alleles do not always correspond.

‡ Identical alleles were found in the Waorani by Watkins *et al.*⁸.

§ Identical B48 alleles were obtained from Piman Indian cells 00136 (HLA-A2: B48,BN21: Cw3) and 02627 (HLA-A2, 24: B5,48: Cw8).

ruled out, it seems less likely that these alleles were all present in the founding populations that colonized America, but have been subsequently lost from or gone undetected in modern Asian and North American Indian populations.

Humans first entered America by the Behring land bridge

Exon 3

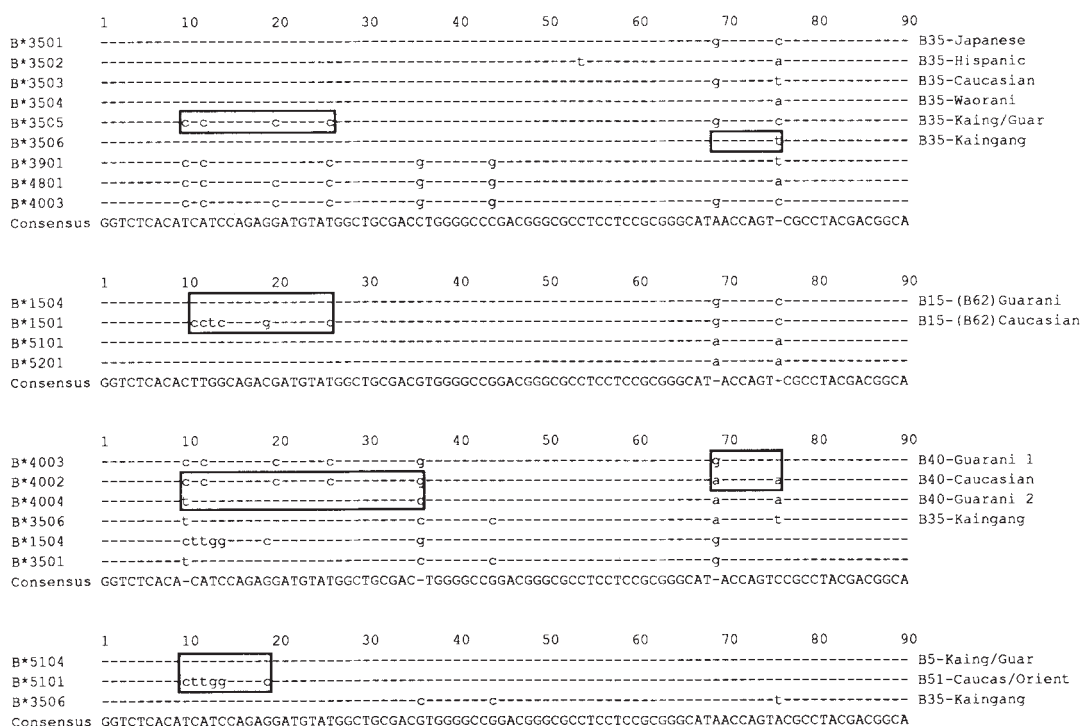


FIG. 2 The localized clusters of nucleotide substitutions in exon 3 that distinguish the *HLA-B35*, *B40*, *B51* and *B62* alleles of the Kaingang and Guarani from those in other populations. These clusters are boxed. The B*3504 subtype found in the Waorani and reported by Watkins *et al.*⁸ is also included.

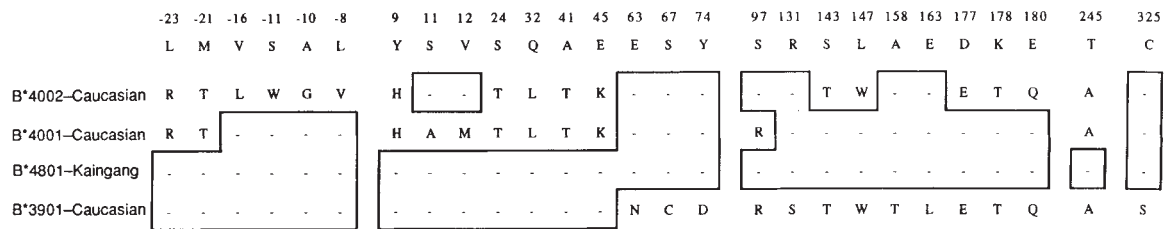


FIG. 3 The structural relationships of B*4801. Positions of amino-acid substitution between B*4801 and B*3901, B*4001 and B*4002 are shown.

Numbering begins at the amino terminus of the mature heavy-chain protein. Identities with B*4801 are boxed.

under conditions of glaciation (reviewed in ref. 6). The descendants of these people subsequently populated much of America and during their southward migrations to temperate and tropical zones would have experienced changing environments with increasing biological and pathogen diversity. Under such circumstances, demands upon antigen presentation by class I HLA molecules are unlikely to have remained constant, and the selection of new *B* alleles in South America could represent responses to the changing environment. Some were, perhaps, selected by specific diseases, as documented in West Africa for B53 and severe malaria¹⁸; others may have been selected for a more general capacity to expand antigen-presenting function against a range of pathogens through increased heterozygosity^{19,20}. For populations having small numbers of alleles and more homozygotes, like Amerindian tribes, the advantage of any new allele is probably greater than in populations with many alleles like caucasians. The action of such selection, which is not dominated by a particular disease, could account for allelic differences between tribes from the same area, like the Kaingang and Guarani. Frequency-dependent effects could also come into play as individuals with low-frequency alleles are more likely to have heterozygote progeny than those with high-frequency alleles²¹.

HLA polymorphism has primarily been studied in large urban communities where many low-frequency alleles are the norm²². On the timescale of human evolution such populations are a recent phenomenon, the products of accumulated merger of previously independent populations. Our study of the South American Indians suggests that the large arrays of *HLA* alleles in many modern populations may not be the result of selection but of mixing many smaller communities, each with a small number of distinctive alleles at relatively high frequency. □

New recombinant *HLA-B* alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci

David I. Watkins, Stephen N. McAdam, Xiaomin Liu, Clarice R. Strang*, Edgar L. Milford†, Cindy G. Levine, Theodore L. Garber‡, Alex L. Dogon, Carol I. Lord, Steven H. Ghim, Gary M. Troup§, Austin L. Hughes|| & Norman L. Letvin

Harvard Medical School, New England Regional Primate Research Center, One Pine Hill Drive, Southborough, Massachusetts 01772, USA

* PO Box 9282, Guayaquil, Ecuador, South America

† Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA

‡ Texas A&M University, College of Veterinary Medicine, College Station, Texas 77843, USA

§ Department of Pathology, School of Medicine, University of New Mexico, New Mexico 87131, USA

|| Department of Biology, 208 Mueller Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802, USA

EVIDENCE suggests that the New World was colonized only 11,000–40,000 years ago by Palaeo-Indians¹. The descendants of these Palaeo-Indians therefore provide a unique opportunity to study the effects of selection on major histocompatibility complex class I genes over a short period. Here we analyse the class I alleles of the Waorani of South America and the Zuni of North America. Four of the Waorani *HLA-B* alleles were new functional variants which could be accounted for by intralocus recombination. In contrast, all of the Zuni *HLA-A* and *-B* molecules were present in caucasians and orientals. This suggests that the new Waorani *HLA-B* variants arose in South America. The description of four new *HLA-B* alleles in the Waorani and another five new *HLA-B* alleles from two other tribes of South American Amerindians² indicates that the *HLA-B* locus can evolve rapidly in isolated populations. These studies underline the importance of gathering genetic data on endangered native human populations³.

The Waorani (also known as the Auca) have been one of the most isolated of South American tribes and thus are unlikely to have any admixture of European major histocompatibility complex (MHC) class I alleles. This proto-agricultural tribe, which now numbers approximately 600 individuals, appears to be highly inbred and quite distinct, both genetically and linguistically, from other South American tribes⁴. Serological analysis demonstrated that the Waorani expressed many of the typical Amerindian MHC class I alleles (*HLA-A2*, *-A24*, *-A31*, *-B15*, *-B35*, *-B39*, *-B40*, *-B51*, *-Cw1*, *-Cw4* and *-Cw7*). But some of the alloantisera used for typing showed unusual reactivities with the Waorani lymphocytes. In fact, one-dimensional isoelectric focusing analysis demonstrated that these alloantisera were recognizing two different molecules in the Waorani (Fig. 1 and

Received 7 February; accepted 2 April 1992.

- Salzano, F. M. *et al.* *Am. J. Phys. Anthropol.* **53**, 513–524 (1980).
- Neves, W. A., Salzano, F. M. & Da Rocha, F. J. *Am. J. Phys. Anthropol.* **67**, 13–17 (1985).
- Kostyu, D. D. & Amos, D. B. *Tissue Antigens* **16**, 111–123 (1981).
- Black, F. L. *Am. J. Hum. Genet.* **36**, 1318–1331 (1984).
- Cavalli-Sforza, L. L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 6002–6006 (1988).
- Fagan, B. M. *The Great Journey: The Peopling of Ancient America* (Thames and Hudson, London, 1987).
- Zemmour, J. & Parham, P. *Immunogenetics* **33**, 310–320 (1991).
- Watkins, D. I. *et al.* *Nature* **357**, 329–333 (1992).
- The Immunobiology of HLA1* (ed. Dupont, B.) (Springer-Verlag, New York, 1989).
- Bodmer, J. G. *et al.* *Tissue Antigens* (in the press).
- Williams, R. C. & McAuley, J. *Hum. Immun.* **33**, 39–46 (1992).
- Salter, R. D. *et al.* *Nature* **338**, 345–347 (1989).
- Castañón, A. R. & López de Castro, J. A. *Immunogen.* **34**, 281–285 (1991).
- Zemmour, J. *et al.* *Tissue Antigens* (in the press).
- Neel, J. V. *Am. J. Hum. Genet.* **30**, 465–490 (1978).
- Klein, J. *Natural History of the Major Histocompatibility Complex* 627–633 (Wiley, New York, 1986).
- Parham, P. *et al.* *J. Immun.* **142**, 3937–3950 (1989).
- Hill, A. V. S. *et al.* *Nature* **352**, 595–600 (1991).
- Hughes, A. L. & Nei, M. *Nature* **335**, 167–170 (1988).
- Hedrick, P. W., Whittam, T. S. & Parham, P. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5897–5901 (1991).
- Wills, C. *Immunol. Rev.* **124**, 165–220 (1991).
- Hedrick, P. W. & Thomson, G. *Genetics* **104**, 449–456 (1983).
- Ennis, P. D. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **87**, 2833–2837 (1990).
- Zemmour, J. *et al.* *J. Immun.* **148**, 1941–1948 (1992).

ACKNOWLEDGEMENTS. We thank D. Watkins and co-workers for sharing their data on the class I alleles of the Waorani and for exchanging manuscripts before publication; members of the Guarani and Kaingang tribes for their cooperation; W. C. Knowler for Epstein-Barr virus-transformed B-cell lines from members of the Gila River Indian community; and V. Ogbuehi for preparing the manuscript. This research was supported by grants from the USPHS, the Cancer Research Institute and the Leukemia Society of America.

TABLE 1 Segmental exchange has generated new *HLA* alleles in the Waorani

Allele designation	Serological reactivity	Exchanged segment	Potential donors
A*0201	A2	—	—
A*0204	A2	Exon 3; 18	A*31012, A*2402
A*0211	A2	Exon 2; 216–218	A*31012
A*2402	A24	—	—
A*31012	A31	—	—
B*1504	B15	Exon 3; 10–25	B*52012
B*3504	B35	Exon 3; 68–75	B*4801, B*52012
B*3903	B39	Exon 3; 19	B*4801
B*4801	B40	—	—
B*4802	B35†	Exons 1 and 2; 1–198	B*3501‡
B*52012	B51	Exon 2; 67	B*51012‡
C*0102	Cw1	—	—
C*0401	Cw4	—	—
C*X	Cw7	—	—

† Only two of four B35-specific alloantisera reacted with B*4802.

‡ Alleles present in the Zuni.

Table 1). We cloned and sequenced full-length complementary DNAs encoding these Waorani MHC class I molecules to determine the molecular basis for the unusual reactivities of the alloantisera (Fig. 2). Surprisingly, four of the Waorani *HLA-B* alleles were new functional variants. All of these new Waorani *HLA-B* alleles could have been created by recombination of DNA segments exchanged between different alleles of the same locus (Table 1). This would suggest that this type of recombination can be an important mechanism for creating new alleles at the MHC class I loci⁵. Thus, for example, the Waorani *HLA-B*1504* allele differs from *HLA-B*1501* by six bases over a segment of 16 bases; these six substitutions are found in the

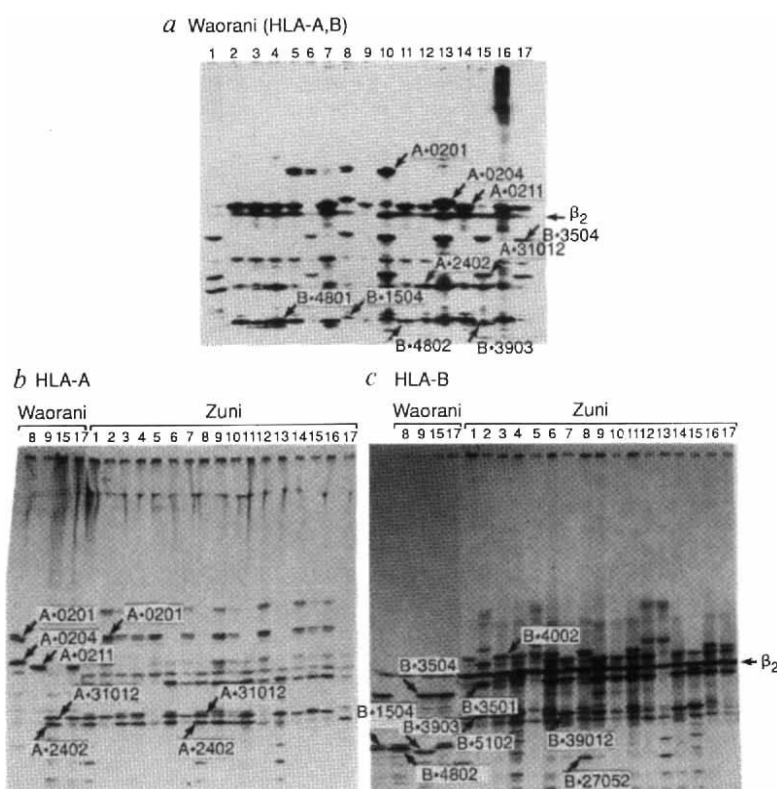
Waorani *HLA-B*52012*, a potential donor for this segment. The *HLA-B*4802* allele is a recombinant between *HLA-B*4801* and *HLA-B*3501* (Fig. 2 and Table 1). In contrast to the unique Waorani *HLA-B* alleles, the Waorani *HLA-A* and *-C* alleles were unremarkable and all but one had previously been described in caucasian and oriental populations (Fig. 2).

To determine whether the new Waorani *HLA-B* variants arose in South America or whether they were actually rare oriental variants, we next analysed MHC class I alleles from another Amerindian tribe derived from the founding Palaeo-Indians. Like the Waorani, the Zuni are descendants of the Palaeo-Indians and a high frequency of albinism (1/160) indicates that these Native Americans are also highly inbred^{6,7}. About 7,000 Zuni inhabit a 440,000-acre reserve in West Central New Mexico. Serological analysis indicated that the Zuni expressed only a limited number of MHC class I alleles (*HLA-A2*, *-A24*, *-A31*, *-B15*, *-B27*, *-B35*, *-B39*, *-B40*, *-B48*, *-B51*, *-Cw1*, *-Cw3*, *-Cw4* and *-Cw7*). Furthermore, most of these alleles were shared with other Amerindian tribes including the Waorani. Indeed, one-dimensional isoelectric focusing and sequence analysis demonstrated that the three Zuni *HLA-A* alleles (*HLA-A*0201*, *HLA-A*2402* and *HLA-A*31012*) were identical to those expressed by the Waorani (Figs 1b and 2). In marked contrast to the constancy of the *HLA-A* locus, however, similar analysis of the Zuni *HLA-B* alleles revealed that only one of them (*HLA-B*4801*) was found in the Waorani (Figs 1c, and 2). Remarkably, all of these Zuni *HLA-B* alleles have been sequenced previously from caucasians and orientals (Fig. 2). Thus, no new functional variant *HLA* alleles were present in this tribe of North American Amerindians.

The isolation of four new functional *HLA-B* alleles from the Waorani and another five new *HLA-B* alleles from the Kaingang and Guarani tribes of Southern Brazil² suggests that the diversification of the human MHC class I loci may actually take place much more rapidly than previously estimated^{8,9}. Two

FIG. 1 a, One-dimensional isoelectric focusing (IEF) analysis of the MHC class I molecules of the Waorani. The monoclonal antibody W6/32 (ref. 18) was used to precipitate *HLA-A* and *-B* alleles from Waorani lymphocytes. (W6/32 precipitates *HLA-C* alleles poorly and thus only *HLA-A* and *-B* molecules are precipitated strongly.) Each strong band on the IEF gel represents an allele of the *HLA-A* or *-B* locus. Allelic designations of these bands are indicated where possible (based on serological typing and one-dimensional IEF analysis using *HLA-A*- (F4-72) and *HLA-B*-specific (4E) monoclonal antibodies; data not shown). Several of the Waorani MHC class I molecules were incorrectly recognized by the alloantisera used. Individual number 11 shows an unusual alloantiserum reactivity. He is typed as *HLA-B35* yet his lymphocytes do not express the Waorani *HLA-B*3504* (shown just below β_2 -microglobulin; see individual 17). But another band representing the molecule recognized by the *HLA-B35*-specific alloantisera in individual 11 is present at the bottom of the gel. This molecule (*HLA-B*4802*) is a chimera between *HLA-B*3501* and *HLA-B*4801*, explaining the crossreactivity with the *HLA-B35*-specific antisera. b and c, One-dimensional IEF analysis of the MHC class I molecules of the Zuni. The monoclonal antibodies F4-72 (b; *HLA-A*) and 4E (c; *HLA-B*) were used to precipitate preferentially *HLA-A* and *-B* alleles, respectively, from Waorani and Zuni lymphocytes. Bands on the IEF gels are indicated where possible. The Zuni *HLA-A* molecules were identical to the Waorani *HLA-A* molecules. The Zuni *HLA-B* molecules, however, focused differently from the *HLA-B* molecules of the Waorani, indicating that they had different charges.

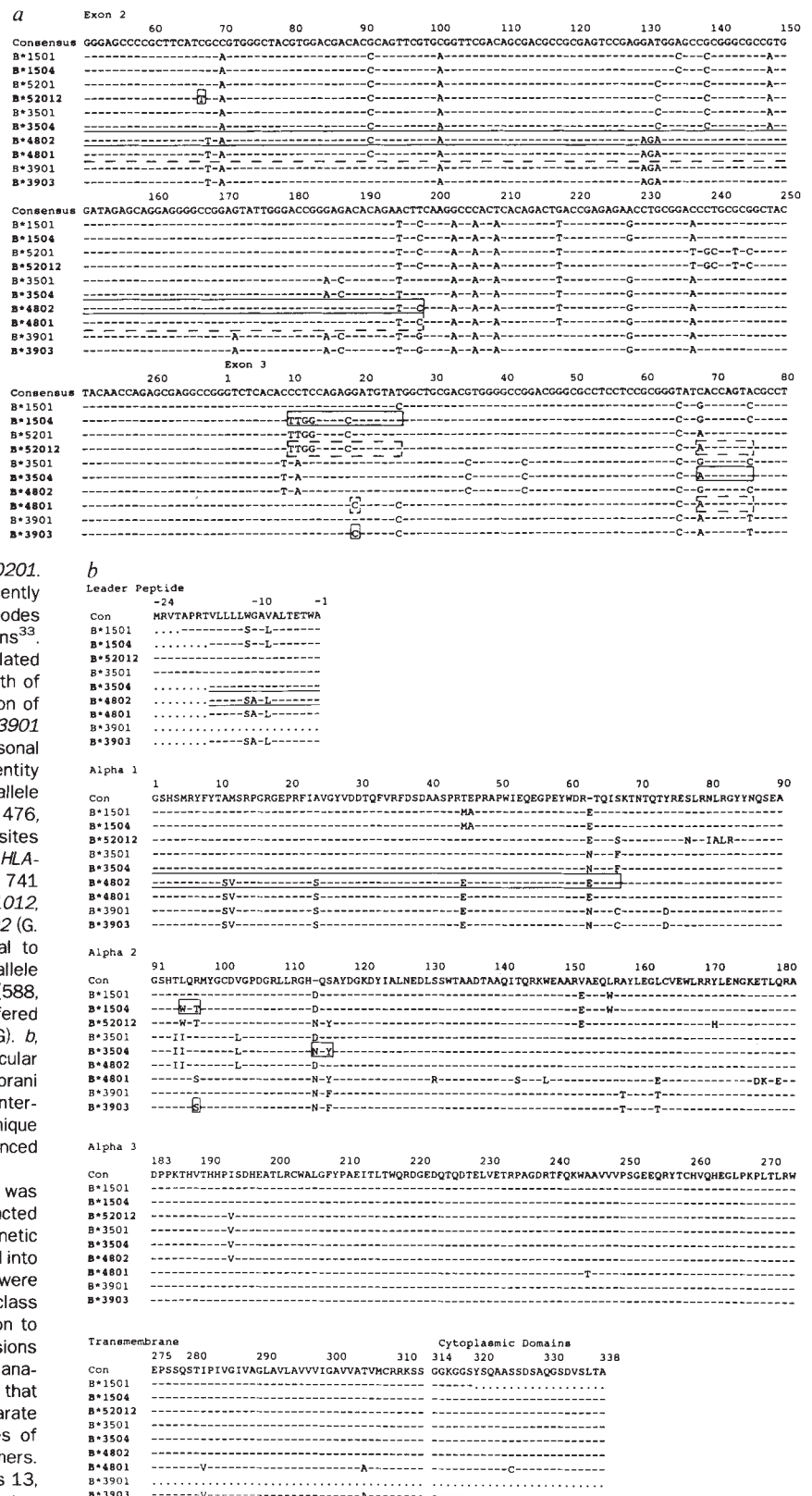
METHODS. Blood samples were taken from 17 Waorani in Ecuador and shipped on ice to Boston. Lymphocytes were separated using Ficoll/diatrizoate and both phytohaemagglutinin-stimulated T-cell lines and Epstein-Barr virus (EBV)-transformed B-cell lines were generated. EBV-transformed lymphocytes were radiolabelled and MHC class I molecules precipitated and subjected to IEF as before¹⁹. Phytohaemagglutinin-stimulated lymphocytes or EBV-transformed cell lines were serotyped using the lymphocytotoxicity assay



(NIH standard). Similarly, EBV-transformed lymphocyte lines were generated from the peripheral blood lymphocytes of 17 Zuni and their MHC class I molecules were radiolabelled, precipitated and subjected to one-dimensional IEF.

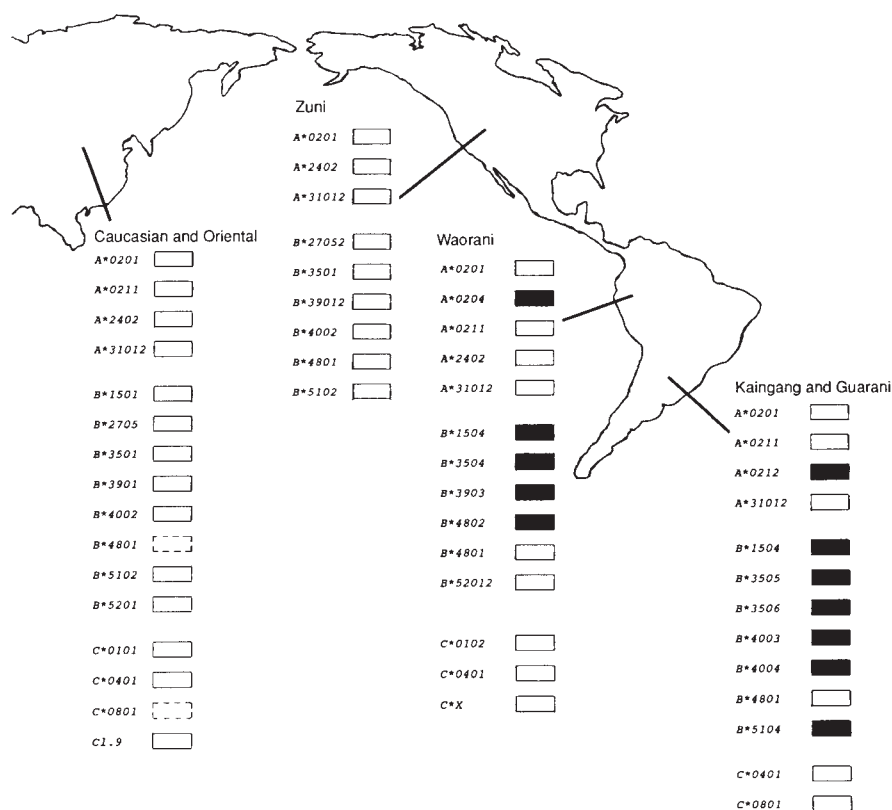
FIG. 2 *a*, All of the Waorani *HLA-B* molecules are new molecular variants. Three hundred bases of the human consensus sequence⁵ are shown and the Waorani alleles (in bold) and their previously sequenced counterparts are compared with it. Segments of DNA that distinguish the Waorani alleles from previously sequenced alleles are boxed with solid lines. Potential donor sequences are boxed with broken lines. The Waorani alleles and the Zuni alleles were compared with the following alleles: *HLA-A*0201* (ref. 20), -*A*2402* (ref. 21), -*A*3101* (ref. 22), -*B*1501* (ref. 23), -*B*2705* (ref. 24), -*B*3501* (ref. 25), -*B*3901* (ref. 26), -*B*4001* (ref. 27), -*B*4002* (ref. 27), -*B5101* (ref. 28), -*B*5201* (ref. 28), -*C*0101* (ref. 29), -*C*0202* (ref. 30), -*C*0401* (J. Zemmour and P. Parham, personal communication) and clone 9 (ref. 31). The Waorani *HLA-A*0201*, *HLA-A*2402* and *HLA-C*0401* alleles were identical to previously sequenced *HLA* alleles. The Waorani *HLA-A*31012* and *HLA-B*52012* variants each differed from the previously sequenced *HLA-A*31011* and *HLA-B*5201* alleles by only one synonymous substitution. The two Waorani *HLA-A2* variants, *HLA-A*0204* (362, G to T substitution) and *HLA-A*0211* (290, C to T and 292, C to G substitutions) differed by one and two non-synonymous nucleotide substitutions respectively from *HLA-A*0201*. cDNA identical to the Waorani *HLA-A*0204* allele has recently been isolated from an Amerindian cell line³². This cDNA encodes an IEF variant that has only been described in Amerindians³³. The Waorani *HLA-A*0211* allele has also been recently isolated from a cell line derived from an oriental individual³⁴. Both of these *HLA-A2* variants may have arisen by recombination of *HLA-A*0201* with *HLA-A*31012* (Table 1). The *HLA-B*3901* allele has recently been re-sequenced (M. Takiguchi, personal communication) and the corrected sequence is shown. Identity is indicated with a dash. The Waorani *HLA-B*3903* allele differed at two sites from *HLA-B*3901* (363, G to C and 476, C to G). The Waorani *HLA-C*0102* allele differed at two sites from *HLA-C*0101* (605 G to C, 606 C to G). The Waorani *HLA-C*X* allele differed at two sites from clone 9 (480 C to G, 741 G to C). The Zuni *HLA-A*0201*, *HLA-A*2402*, *HLA-A*31012*, *HLA-B*3501*, *HLA-B*4002*, *HLA-B*4801* and *HLA-B*5102* (G. Kawaguchi, manuscript submitted) alleles were identical to previously sequenced *HLA* alleles. The Zuni *HLA-B*27052* allele differed at three synonymous sites from *HLA-B*2705* (588, G to T, 690, A to G, 780, G to A). The Zuni *HLA-B*39012* differed at one synonymous site from *HLA-B*3901* (476, C to G). *b*, Predicted amino-acid sequences of the new Waorani molecular variants. A consensus sequence is shown and the Waorani *HLA-B* alleles (in bold) and their previously sequenced counterparts are compared; identity is indicated with a dash. Any unique amino-acid substitutions not found in previously sequenced alleles are indicated by boxing them.

METHODS. Polymerase chain reaction (PCR) from cDNA was carried out as previously described³⁵. Briefly, RNA was extracted from $2-5 \times 10^6$ lymphocytes using oligo(dT) bound to magnetic beads. After cDNA synthesis and PCR, inserts were ligated into pSP65 and at least three full-length copies of each cDNA were sequenced to avoid PCR errors³⁶. At least four MHC class I-specific sequencing primers were used in each direction to generate double-stranded sequence and any compressions were resolved using dITP. Sequences were digitized and analysed using software from IBI. To rule out the possibility that the recombinant cDNAs were PCR artefacts, three separate PCR amplifications were carried out on the lymphocytes of seven different Waorani using three combinations of primers. In the first instance the MHC class I cDNAs of individuals 13, 14 and 15 were amplified from lectin-stimulated lymphocytes using primers LPHIII and 3'UTRI³⁷. *HLA-A*-specific amplification was then carried out on EBV-transformed lymphocytes from individual 10 using 5'ASP (5'-GCAAGCTTATGCGCAT-3') and 3'ASP (5'-GCGAATTCGTCTCACATTTAC-3'). *HLA-B* and *-C* cDNAs were amplified from the EBV-transformed lymphocytes of individual numbers 1, 4, 11 and 14 using primers LPHIII and MM2 (5'-GCGAATTCAGGCTTTT-3'). cDNAs were cloned from the following individuals; *HLA-B*3504*, *HLA-B*52012* and *HLA-C*X* (number 1); *HLA-B*4802* and *HLA-B*4801* (no. 4); *HLA-A*0201* (no. 10); *HLA-B*4802* (no. 11); *HLA-A*0204*, *HLA-A*2402*, *HLA-B*3504*, *HLA-B*4801* and *HLA-C*0401* (no. 13); *HLA-A*0211*, *HLA-B*1504*, *HLA-B*4802*, *HLA-*



*C*0102* and *HLA-C*0401* (no. 14); *HLA-A*31012*, *HLA-A*2402*, *HLA-B*3903* and *HLA-C*0401* (no. 15). *HLA-A*-specific amplification was carried out on Zuni EBV-transformed lymphocytes using 5'ASP and 3'ASP. *HLA-B* cDNAs were amplified from the Zuni EBV-transformed lymphocytes using LPHIII and MM2. cDNAs were cloned from the following individuals; *HLA-B*3501*, *HLA-B*5102* and *HLA-A*0201* (no. 2); *HLA-B*4002* and *HLA-B*4801* (no. 5); *HLA-A*2402*, *HLA-A*31012*, *HLA-B*3501* and *HLA-B*4002* (no. 6); *HLA-B*3501*, *HLA-B*39012* (no. 7); *HLA-B*27052* and *HLA-B*4002* (no. 8).

FIG. 3 Distribution of *HLA-A*, *-B* and *-C* alleles in Asia, North America and South America. Previously sequenced caucasian and oriental alleles are indicated by open boxes. New functional variants are shown with closed boxes. Caucasian and oriental alleles that have only been detected using serological techniques and have not been sequenced yet are boxed with broken lines. The Kaingang and Guarani alleles are from ref. 2. At the *HLA-B* locus all six alleles sequenced from the Zuni are known from caucasians and orientals. By contrast, in the three South American tribes studied by us and by Belich *et al.*², only two *HLA-B* alleles are known from orientals, whereas nine are undescribed. This difference between the Zuni and the South American tribes in proportion of previously undescribed alleles is statistically significant (Fisher's exact test; $P < 0.001$).



alternative hypotheses might explain the finding of nine new *HLA-B* variants in the South American Amerindian tribes, whereas only caucasian and oriental alleles were present in the North American Amerindians (Fig. 4). The first hypothesis is that the founding Palaeo-Indian population may have contained all of the new variant alleles of the Ecuadorian and Brazilian Amerindians and the caucasian and oriental alleles of the Zuni. The caucasian and oriental alleles could have then increased in frequency in the Zuni and the new variant alleles could have increased in frequency in the South American tribes. But it is hard to explain how such a partitioning of previously sequenced and variant alleles between the Zuni and the South American Amerindians, respectively, could occur by chance. The second hypothesis to explain these findings is that the caucasian and oriental alleles of the Zuni represent the MHC class I alleles of the founding Palaeo-Indians. On entering the tropics the Palaeo-Indians could have encountered new selective pressures resulting in any new variants reaching appreciable frequencies in the South American Amerindians. Additionally, a limited number of founding MHC haplotypes would be consistent with both the paucity of mitochondrial haplotypes present in the Amerindians^{10,11} and with the small number of *HLA-A* alleles found in all of the descendants of the Palaeo-Indians¹².

The presence of new variant *HLA-B* alleles only in the South American tribes would suggest that both selection and genetic drift may have been involved in increasing the frequency of these alleles. It has recently been suggested that *HLA-Bw53* may confer a survival advantage on individuals recovering from acute malarial infection¹³. Other MHC alleles may play a similar part in immune responses to other pathogens. On entering the American tropics, Palaeo-Indian populations may have encountered several new pathogens (such as *Trypanosoma cruzi* and *Leishmania braziliensis*). In these circumstances, overdominant selection¹⁴ might cause new recombinant alleles to increase in frequency. Additionally, population densities in South America are low and there is very little intermarriage between linguistic groups, serving to keep small populations reproductively isolated¹⁵. New variant alleles would increase in frequency in such populations. Thus, the high proportion of *HLA-B* recombinants

in the South American Amerindian population is consistent with the idea that small isolated populations can be a potential source of evolutionary diversity¹⁶.

The unusual MHC class I alleles of the Waorani, Kaingang and Guarani could have important implications for tracing the ancient migrations of human populations. These data raise the intriguing possibility that some MHC class I variants may indeed be restricted to certain races. *HLA-B27* variants show a restricted racial distribution¹⁷, with *HLA-B*2706* being found only in orientals and *HLA-B*2703* in African and American blacks. Likewise, some of the MHC class I diversity found in Amerindians may have taken place after the arrival of the Palaeo-Indians in South America. Therefore it may be possible to trace human migrations using MHC class I cDNA sequence analysis. □

Received 7 February; accepted 6 April 1992.

1. Marshall, E. *Science* **249**, 738–741 (1990).
2. Belich, M. P. *et al. Nature* **357**, 326–329 (1992).
3. Roberts, L. *Science* **252**, 1614–1617 (1991).
4. Larrick, J. W. *et al. Am. J. Phys. Anthropol.* **66**, 445–453 (1985).
5. Parham, P., Lawlor, D. A., Lomen, C. E. & Ennis, P. D. *J. Immunol.* **142**, 3937–3950 (1989).
6. Woodbury, R. B. in *Handbook of North American Indians* Vol. 9 (ed. Ortiz, A.) 467 (Smithsonian Institution, Washington DC, 1979).
7. Troup, G. M., Harvey, R. L., Walford, R. L., Smith, G. S. & Sturgeon, P. in *Histocompatibility Testing 1972* (eds Dausset, J. & Colombani, J.) 339–344 (1973).
8. Lawlor, D. A., Ward, F. E., Ennis, P. D., Jackson, A. P. & Parham, P. *Nature* **335**, 268–271 (1988).
9. Mayer, W. E. *et al. EMBO J.* **7**, 2765–2774 (1988).
10. Schurr, T. G. *et al. Am. J. Hum. Genet.* **46**, 613–623 (1990).
11. Torroni, A. *et al. Genetics* **130**, 153–162 (1992).
12. Black, F. L., Pandey, J. P. & Santos, S. E. B. in *Origens, Adaptações e Diversidade Biológica do Homem Nativo da Amazonia* (eds Neves, W. & Nassif, R.) (Museu Paraense, E. Goeldi, Belém, 1990).
13. Hill, A. V. S. *et al. Nature* **352**, 595–600 (1991).
14. Hughes, A. L. & Nei, M. *Nature* **335**, 167–170 (1988).
15. Salzano, F. M. & Cavigliari-Jaques, S. M. *South American Indians. A Case Study in Evolution* (Clarendon, Oxford, 1988).
16. Wright, S. *Science* **87**, 430–431 (1938).
17. López de Castro, J. A. *Immun. Today* **10**, 239–246 (1989).
18. Parham, P., Barnstable, C. J. & Bodmer, W. F. *J. Immunol.* **123**, 342–349 (1979).
19. Watkins, D. I., Kannagi, M., Stone, M. E. & Letvin, N. L. *Eur. J. Immunol.* **18**, 1425–1432 (1988).
20. Koller, B. H. & Orr, H. T. *J. Immunol.* **134**, 2727–2733 (1985).
21. Little, A.-M., Madrigal, J. A. & Parham, P. *Immunogenetics* **35**, 41–45 (1991).
22. Kato, K., Traponi, J., Dupont, B. & Yang, S. Y. *J. Immunol.* **143**, 3371–3378 (1989).
23. Pohla, H., Kuon, W., Tabaczewski, P., Doerner, C. & Weiss, E. H. *Immunogenetics* **29**, 297–307 (1989).
24. Seeman, G. H. A., Rein, R. S., Brown, C. S. & Ploegh, H. L. *EMBO J.* **5**, 547–552 (1986).
25. Ooba, T. *et al. Immunogenetics* **30**, 76–80 (1989).

26. Muller, C. A. *et al. Immunogenetics* **30**, 200–207 (1989).
27. Ways, J. P., Lawlor, D. A., Wan, A. M. & Parham, P. *Immunogenetics* **25**, 323–328 (1987).
28. Hayashi, H. *et al. J. Immun.* **142**, 306–311 (1989).
29. Gussow, D. *et al. Immunogenetics* **25**, 313–322 (1987).
30. Parham, P. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 4005–4009 (1988).
31. Cianetti, L. *et al. Immunogenetics* **29**, 80–91 (1989).
32. Castaño, A. R. & López de Castro, J. A. *Immunogenetics* **34**, 281–285 (1991).
33. Yang, S. Y. in *Histocompatibility Testing 1987* (ed. Dupont, B.) 309–331 (Springer, New York, 1989).
34. Castaño, A. R. & López de Castro, J. A. *Immunogenetics* **35**, 344–346 (1992).
35. Newman, P. J. *et al. J. Clin. Invest.* **82**, 739–743 (1988).
36. Ennis, P. D., Zemmour, J., Salter, R. D. & Parham, P. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2833–2837 (1990).
37. Watkins, D. I. *et al. Immunogenetics* **33**, 79–89 (1991).

ACKNOWLEDGEMENTS. We thank the Waorani and Zuni for blood samples; S. and J. Padilla, I. Gianetti and SAETA Airlines for their help in obtaining lymphocytes; J. A. López de Castro, F. L. Black and M. D. Miller for reviewing the manuscript; S. Y. Yang for monoclonal antibodies F4/72 and 4E; and M. P. Belich, J. A. Madrigal, W. H. Hildebrand, R. C. Williams, R. Luz, M. L. Petzl-Erler and P. Parham for reviews of their manuscript². This work was supported by grants from the NIH.

Molecular cloning of the receptor for human antidiuretic hormone

Mariel Birnbaumer, Anita Seibold, Stephanie Gilbert, Masami Ishido*, Claude Barberis*, Anaid Antaramian, Philippe Brabet & Walter Rosenthal

Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030, USA

ANTIURESIS, the recovery of water from the lumen of the renal collecting tubule, is regulated by the hypothalamic release of antidiuretic hormone (ADH), which binds to specific receptors on renal collecting tubule cells, stimulates adenylyl cyclase and promotes the cyclic AMP-mediated incorporation of water pores into the luminal surface of these cells^{1–3}. We report here the isolation of the human ADH receptor gene using a genomic expression cloning approach⁴. The gene was used to clone the complementary DNA from a human renal library. The deduced amino-acid sequence of the receptor yields a hydropathy profile characteristic of receptors with seven putative transmembrane regions. This and the comparison with other cloned receptors indicates that the ADH receptor is a member of the superfamily of G-protein-coupled receptors.

We constructed genomic libraries with DNA from the tertiary transformant cell HTB-3 both in the cosmid vector pWE16 (ref. 5) and in the phage λ EMBL3 (ref. 6), and screened for clones positive for human repetitive DNA. The cloned cosmids and phages were then tested in stable transfection assays for their ability to confer ADH or arginine vasopressin (AVP) responsiveness to L cell adenylyl cyclase. One of the λ EMBL3 phages (clone number 14; see Table 1 and Fig. 1a) proved to be positive in this screening. Several of the new cell clones (LV2 cells) were expanded and tested for specific [³H]AVP binding using an assay that has a limit of detection of 500 sites per cell. In the G9 cell, specific [³H]AVP binding detected about 200,000 sites per cell (Fig. 2a). Similar levels of [³H]AVP binding were found in two other cell lines isolated in this way, the D4 and E7 cells.

To locate the ADH receptor gene sequence in λ EMBL3–14, *Bam*HI restriction fragments (Fig. 1a), were cloned into the plasmid Bluescript KS(+) and DNA from the individual plasmids was cotransfected with the *tk* selection gene into naïve *Ltk*[−] cells by the calcium phosphate method. Stable transformants selected in HAT (hypoxanthine/aminopterin/thymidine) medium were tested for the acquisition of AVP-stimulated adenylyl cyclase activity. Of 72 *tk*⁺ cell lines isolated from transfections with Bluescript containing a 2.2 kilobase (kb) *Bam*HI restriction fragment (pBS-B2.2), 11 contained the vasopressin receptor

transcription unit as shown by *in situ* adenylyl cyclase assays⁴. Several of these cell lines were expanded and their vasopressin responsiveness confirmed (Table 1). Two of these LV2 cell lines, clones E2 and H4, were analysed further for number of AVP-binding sites and responsiveness of its adenylyl cyclase to AVP and analogues. In contrast to the λ EMBL3–14-derived LV2.D4 cell, the pBS-B2.2-derived LV2.E2 and H4 cells expressed only about 5,000 sites per cell (not shown). Their adenylyl cyclase activity was still stimulated 6–8-fold by vasopressin (shown for the E2 cell in Fig. 2b). In agreement with previous studies with the HTB-2 cell⁴ the rank order of potency with which vasopressin and its analogues stimulated the pBS-B2.2-transformed cells continued being AVP > dDAVP $\gg$ LVP $\gg$ OT (dDAVP, deamino-D-arginine vasopressin; LVP, lysine vasopressin; OT, oxytocin; Fig. 2b). This is the rank order of potency of the receptor through which AVP stimulates the renal G_s/adenylyl cyclase complex⁷ and indicates that the receptor encoded in the 2.2-kb *Bam*HI fragment of λ EMBL3–14 was the type-2 vasopressin receptor (V2R) as defined by Michell *et al.*⁸. The type-1 receptor (V1R) has a different coupling specificity. It stimulates the G_p/phospholipase C complex instead of the G_s/adenylyl cyclase complex and interacts with the above peptides with a rank order of potency AVP $\gg$ LVP $\gg$ OT $\gg$ dDAVP⁹.

Using the 2.2-kb *Bam*HI fragment as probe, we detected a prominent messenger RNA of about 1.9 kb in V2R-expressing D4, G9 and E7 cells. A specific oligo(dT) primed complementary DNA library was constructed in λ gt10 using as template poly(A)⁺ RNA from the G9 cell. The 2.2-kb *Bam*HI DNA fragment was used to screen for recombinant phages from this library. One contained a 1.7 kb insert (V2R-G9.3) that was cloned into Bluescript KS(+) to give pBS-G9.3. The G9.3 insert was sequenced in both directions *in situ* using the dideoxy chain termination method. It was 1,663 nucleotides (nt) long, and contained an open reading frame of 1,113 nt that was spanned by 67 nt at its 5' end and 456 nt at its 3' end, of which the last 27 constituted a poly(A)⁺ track.

The identity of the cDNA cloned from the G9 cell was demonstrated by two means: (1) the G9 cell cDNA was subcloned into the expression plasmid pKNH (ref. 10) and the resulting plasmid, pK-V2R-G9.3 was shown to direct the insert-dependent expression of AVP-stimulated adenylyl cyclase in COS.M6 cells (Fig. 2c); and (2) a new cDNA, V2R-hk.5, was cloned from a human kidney cDNA library using two oligonucleotides based on the sequence of the G9.3 clone as screening probes. The nucleotide composition of V2R-hk.5 cDNA differed from that derived from the G9 cell only in that it was four nucleotides longer at its 5' end (Fig. 1d). Northern blot analysis using the 1.7 kb human kidney cDNA as a probe detected a 1.9–2.0-kb RNA in kidney RNA from two species: rat and rabbit, as well as a smaller transcript in the porcine kidney cell LLCPK₁ known to express a V2 lysine vasopressin receptor coupled to adenylyl cyclase. No signal was detected in four other rat tissues, except for a faint signal in liver corresponding to a transcript of 1.8 kb (Fig. 3). This signal could be due to hybridization with the V1R or to extrarenal V2R.

Hydropathicity analysis done according to Kyte and Doolittle^{11,12}, predicted seven putative transmembrane segments (Fig. 1c). Thus, in agreement with the fact that this receptor activates G_s, its structure conforms to the general structure of a G-protein-coupled receptor. The V2R cDNA predicts a protein of 371 amino acids with a calculated *M_r* of 40,285 in the absence of post-translational modifications.

It is possible to classify the G-protein-coupled receptors based either on a common or similar ligand into subfamilies showing limited sequence identities only in their transmembrane regions (not exceeding 25–35%). In this context, the V2 receptor seems to belong to a new subfamily. When the transmembrane regions of the V2R are compared to those of the bovine substance K (ref. 13), human S12 serotonin¹⁴, human β 1 adrenergic¹⁵, murine LH (ref. 16), and rat M5 muscarinic¹⁷, they are most similar to

* Present addresses: Institute of Applied Biochemistry, University of Tsukuba, Ibaraki-ken 305, Japan (M.I.); Institute CNRS-INSERM de Pharmacologie-Endocrinologie; Rue de la Cardonille, Montpellier CEDEX 34094, France (C.B.).

TABLE 1 Summary of cloning strategy

Donor DNA/RNA with V2R gene/RNA	Cloning medium	Screen/identification	V2R gene/cDNA recovered in cell/plasmid/phage	Ref.
Human genomic DNA	Ltk ⁻	AC assay*	HTB-1 cell	4
HTB-1 DNA	Ltk ⁻	AC assay	HTB-2 cell	4
HTB-2 DNA	Ltk ⁻	AC assay	HTB-3 cell	This report
HTB-3 DNA	λEMBL3	Huma <i>Alu</i> repeats	λEMBL3-14 phage	This report
λEMBL3-14 DNA	Ltk ⁻	AC assay	LV2.D4, .G9 and .E7 cells	This report
2.2 kb <i>Bam</i> HI fragment	Bluescript	Minilysate	pBS-B2.2 plasmid	This report
pBS-B2.2 DNA	Ltk ⁻	AC assay	LV2.E2 and .H4 cells	This report
LV2.G9 poly(A) ⁺ RNA	λgt10	2.2 kb <i>Bam</i> HI fragments	V2R-G9.3 cDNA	This report
Human kidney RNA†	λgt10	V2R-G9.3 oligonucleotides	V2R-hk.5 cDNA	This report

Cotransfection of human genomic DNA and the thymidine kinase gene into murine Ltk⁻ cells generated a primary human genomic expression library in Ltk⁺ stable clones that was screened for the acquisition of any one of 13 adenylyl cyclase-stimulating peptide receptors, and led to the isolation of a stable cell line, called HTB-1, that responded positively to stimulation by AVP. Genomic HTB-1 DNA was used to obtain the secondary transformant HTB-2 that also expressed the human ADH receptor⁴. HTB-2 DNA then was used to obtain the tertiary transformant, HTB-3, using the same procedure. HTB-3 exhibits the same characteristics as HTB-2 but much reduced total content of human repetitive DNA, as determined by the polymerase chain reaction using primers specially designed to detect human sequences in hybrid cell lines¹⁸. To isolate the human gene encoding the AVP-response phenotype, we followed the strategy summarized above.

* Adenylyl cyclase assay.

† Library kindly provided by G. Bell (Univ. Chicago).

FIG. 1 *a*, Restriction map of λEMBL3-14 containing the human V2 vasopressin receptor gene (Fig. 2*a*). Restriction fragments positive for human repetitive DNA are represented as solid boxes; those negative for human, but positive for murine repetitive sequences are represented as open boxes. Enlarged box in λEMBL-314 represents the 2.2 kb *Bam*HI fragment that carries the V2R transcription unit. *b*, Restriction map and sequencing strategy of the receptor V2R-hk.5 cDNA cloned from a human cDNA library in λgt10. The longest open reading frame is represented by an open box. Two additional partial clones were isolated. G9.3-A and G9.3-B, screening oligonucleotides of 45 nt used to clone the hk.5 cDNA. The poly(A)⁺ track at the end of the cDNA begins 12 nt after the AATAAA cleavage and polyadenylation sequence. *c*, Hydropathy profile of the deduced amino-acid sequence of the V2R computed according to Kyte and Doolittle¹² using a window of 15 residues. Regions with hydrophobic indices consistent with formation of a transmembrane spanning segment of 20–25 amino acids are highlighted. *d*, Nucleotide and predicted amino-acid sequences of the human ADH receptor cDNA (clone V2R-hk.5). The sequences of the screening oligonucleotides G9.3-A (antisense) and G9.3-B (sense) used to clone the V2R-hk.5 cDNA are underlined. The ATG of CCACCATGC is assigned as codon 1 on the basis of its close match to the CCA³/CCATGG Kozak consensus sequence for favoured initiation of translation¹⁹ and the presence of an in-frame TGA stop codon 49 nt 5' to it. The amino acids predicted to form transmembrane regions by the Kyte and Doolittle hydrophobicity analysis are boxed. One unique Asn-X-Ser/Thr consensus sequence for glycosylation (Ψ), two Ser/Thr-X-Arg-Lys consensus phosphorylation sites for protein kinase C (▲)²⁰, and one Ser/Thr-X-X-Asp/Glu consensus phosphorylation site for casein kinase II (●)²¹, are highlighted. There is no Arg-Arg/Lys-X-Ser/Thr consensus site for phosphorylation by cAMP-dependent protein kinase²¹. The EMBL accession number is Z11687.

METHODS. Screening for recombinant λEMBL3 phages positive for human repetitive sequences was as follows: 1 × 10⁶ p.f.u. were plated in *Escherichia coli* CES200 at a density of 20,000 p.f.u. per 150 mm dish. Replicas were lifted onto filters embedded with *E. coli* CES200 and phage DNA was amplified⁶ overnight before fixation by NaOH and baking at 68 °C. After pre-hybridization for 3 h at 42 °C in 50% formamide, 5 × SSC, 0.5% SDS, 5 × Denhardt's solution, 50 mM Na-Phosphate buffer, pH 6.8, 100 μg ml⁻¹ sheared salmon sperm DNA, the filters were probed for 20 h at 42 °C in the same medium plus 6 × 10⁶ c.p.m. ml⁻¹ human genomic DNA labelled with ³²P to a specific activity of at least ~10⁹ c.p.m. μg⁻¹ DNA by the random hexanucleotide primed extension method²². Filters were then washed three times with 2 × SSC, 0.5% SDS at room temperature and for 15 min at 62 °C in the same solution, and once in 0.2 × SSC, 0.5% SDS after which the filters were dried and exposed at -70 °C for 20 h to Kodak X-Omat AR films in cassettes fitted with two intensifying screens. Presence of human and murine repetitive DNA in restriction fragments on Southern blots was determined using the same conditions of hybridization and washes as used in the screening of recombinant λEMBL3 phages. The λgt10 libraries were plated in *E. coli* C600 and screened following the same protocol, 1.5 × 10⁶ p.f.u. were screened of the human kidney library.

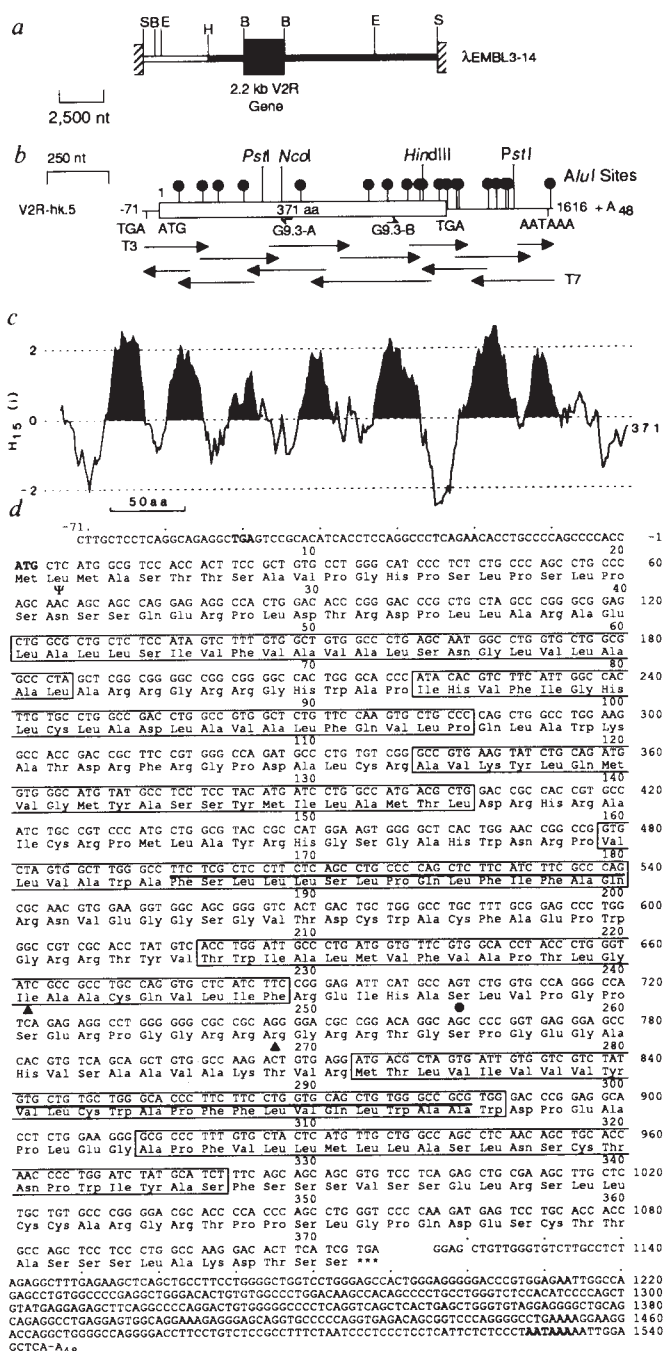
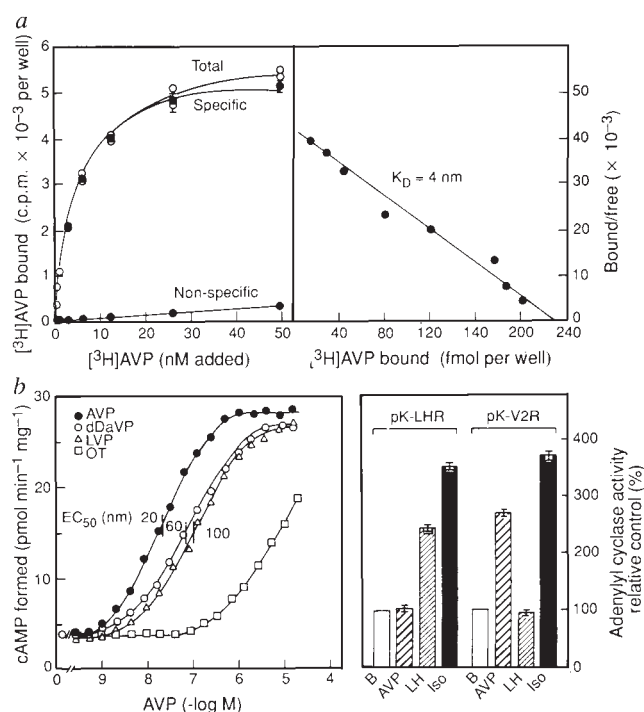


FIG. 2 *a*, Expression of AVP-binding sites in cells transformed with λ EMBL3-14. Left, Binding of [3 H]AVP to intact G9 cells. The binding was specific, neither glucagon nor ACTH (10^{-6} M each) interfered with it, and the primary phenotype, that is, stimulation of adenyl cyclase, was not mimicked by 100 nM of any one of the following peptides: glucagon, adrenocorticotrophic hormone, vasoactive intestinal peptide, secretin, melanotropin, parathyroid hormone, corticotropin-releasing factor or growth hormone releasing factor (not shown). Right, Scatchard transformation of the data. Similar results, that is, an abundance of about 200,000 sites per cell, were obtained with the same cells on a second occasion, as well as twice with LV2R.D4 and LV2R.E7 cells (Table 1). *b*, Expression and dose-response relations for stimulation of adenyl cyclase by the V2 receptor gene encoded in the 2.2 kb *Bam*HI fragment derived from λ EMBL3-14 (Table 1). The LV2.E2 cell line was obtained by stable transformation of an *Ltk*⁻ cell with pBS/KS-V2R-2.2 (Table 1). *c*, Identification of the receptor encoded in the cDNA cloned from the LV2.G9 cell by transient expression in COS.M6 cells (Table 1). Control (no addition of receptor ligand) and isoproterenol-stimulated activities are intrinsic to COS.M6 cells. Similar data were obtained in a repeat experiment. **METHODS.** Binding of [3 H]AVP to cells: LV2.G9 cells (Table 1) grown in six-well plates to a density of 0.75×10^6 cells per well were exposed to the indicated concentrations of [3 H]AVP (68 Ci mmol⁻¹, NEN-DuPont), in the absence (total binding) and the presence of 10 μ M unlabelled AVP (non-specific binding). Reactions were done in duplicate in 0.7 ml Dulbecco's PBS with 2% BSA (D-PBS, 2% BSA), for 2 h at 4 °C. The cells were then washed twice with D-PBS, 2% BSA, and the [3 H]AVP retained by the cells quantified by lysing the cells with 1.0 ml 0.1 M NaOH and determining the radioactivity in a liquid scintillation counter. Determination of adenyl cyclase activity: confluent cells were washed and homogenized, and homogenates were tested for adenyl cyclase activity as previously described⁴, with the additives shown on the figure. AVP, (Arg⁸)-vasopressin; dDAVP, (deamino-Cys⁴, D-Arg⁸)-vasopressin; LVP, (Lys⁸)-vasopressin; OT, oxytocin. Expression of V2R-G9.3 cDNA in COS cells: the 1.7 kb *Eco*RI-*Eco*RI insert of λ gt100-V2R.3 (Table 1) was subcloned by blunt-end ligation into the *Hind*III cloning site of pKNH (ref. 12) to give pK-V2R. COS.M6 cells were seeded in 100-mm dishes at a density of 1×10^6 cells per dish, and transfected one day later with 2.5 μ g supercoiled pK-V2R DNA or pK-LHR DNA (a test pKNH plasmid carrying the murine LH receptor cDNA¹⁶ instead of the cDNA of λ gt10-V2R.3) using a modification of the DEAE dextran procedure of Luthman and Magnusson²³. For each transfection, cells from two 100-mm dishes were collected 36 h after transfection and disrupted with a Dounce homogenizer in 0.5 ml



27%(w/v) sucrose in 1 mM EDTA and 20 mM Na-HEPES, pH 8.0. Aliquots (10 μ l containing 14–18 μ g protein) of these homogenates were then assayed for adenyl cyclase activity in the absence and presence of the additives shown on the figure to determine the identity and specificity of receptor expression. Incubations were in triplicate for 20 min at 32 °C. Control activities of COS.M6 homogenates transfected with the LH receptor cDNA (left) and the V2R-G9.3 cDNA (right) were 30 ± 2 and 26 ± 2 (mean pmol min⁻¹ mg⁻¹ protein $\pm$ s.d.), respectively. When present, AVP, hCG (the LH receptor ligand) and isoproterenol were 100 nM, 5 μ g ml⁻¹ and 1 μ M, respectively.

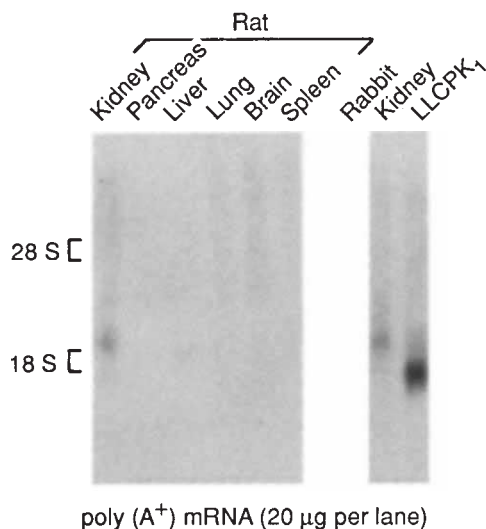


FIG. 3 Analysis of expression of V2R mRNA in rabbit kidney, LLC PK₁ cells, and the rat tissues indicated in the figure.

METHODS. Poly(A)⁺ RNA was isolated from total RNA prepared by the guanidine-isothiocyanate/CsCl method using oligo-dT cellulose chromatography and electrophoresed through a 1.2% agarose gel in the presence of 18 mM MOPS/5 mM Na-acetate/1 mM EDTA buffer, pH 7.0 and 7.4% formaldehyde. The RNA was transferred onto Genescreen-Plus membranes (DuPont). The dry membrane was hybridized overnight with 5×10^6 c.p.m. ml⁻¹ of [32 P] labelled 1.7 kb human cDNA. The blots were washed with $0.2 \times$ SSC, 0.5% SDS for 15 min once at room temperature and once at 55 °C, and autoradiographed for 72 h at -70 °C in the presence of two enhancing screens.

the human β 1 adrenergic receptor (35% sequence identity), and least to the rat muscarinic M5 (25% sequence identity). The ligand-binding properties of the V2 receptor are similar to those of the V1 and the oxytocin receptor. By analogy to other sub-families, the V2 receptor would be expected to be not only closely related pharmacologically but also structurally to the V1 and oxytocin receptors². □

Received 30 December 1991; accepted 31 March 1992.

- Orloff, J. & Handler, J. S. *Am. J. Med.* **42**, 757–768 (1967).
- Jard, S. in *Current Topics in Membranes and Transport* (Vol. 18, (ed. Kleinzeller, A.)) 255–285 (Academic, New York, 1983).
- Harris, H. W., Handler, J. S. & Blumenthal, R. *Am. J. Physiol.* **258**, 237–243 (1990).
- Birnbaumer, M., Hinrichs, V. & Themmen, A. P. N. *Molec. Endocrin.* **4**, 245–254 (1990).
- Wahl, G. M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2160–2164 (1987).
- Kaiser, K. & Murray, N. E. in *DNA Cloning* Vol. 1 (ed. Glover, D. M.) 1–48 (IRL, Oxford, England, 1985).
- Birnbaumer, L., Nakahara, T. & Yang, P.-Ch. *J. biol. Chem.* **249**, 7857–7866 (1974).
- Michell, R. H., Kirk, C. J. & Billah, M. M. *Biochem. Soc. Trans.* **7**, 861–865 (1979).
- Jard, S., Lombard, C., Marie, J. & Devilliers, G. *Am. J. Physiol.* **253**, 41–49 (1987).
- Takeshima, H. et al. *Nature* **339**, 439–445 (1989).
- Lawrence, C. B., Shalom, T. Y. & Honda, S. *Molec. biol. Inform. Resource* (Baylor College of Medicine, Houston, Texas 77030, 1989).
- Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105–132 (1982).
- Masu, Y. et al. *Nature* **329**, 836–838 (1987).
- Levy, F. O. et al. *J. biol. Chem.* (in the press).
- Frielle, T. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7920–7924 (1987).
- Gudermann, T., Birnbaumer, M. & Birnbaumer, L. *J. biol. Chem.* **267**, 4479–4488 (1992).
- Liao, C.-F. et al. *J. biol. Chem.* **264**, 7328–7337 (1989).
- Lichter, P., Ledbetter, S. A., Ledbetter, D. H. & Ward, D. C. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6634–6638 (1990).
- Kozack, M. *J. Cell Biol.* **108**, 229–241 (1989).
- Woodgett, J. R., Gould, K. L. & Hunter, T. *Eur. J. Biochem.* **161**, 177–184 (1986).
- Kennelly, P. J. & Krebs, E. G. *J. biol. Chem.* **266**, 1555–1558 (1991).
- Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6–13 (1983).
- Luthman, H. & Magnusson, G. *Nucleic Acids Res.* **11**, 1295–1308 (1983).

ACKNOWLEDGEMENTS. This research was supported in part by grants from the NIH to M.B., by an INSERM postdoctoral fellowship to P.B. and by a DFG-Heisenberg award to W.R.

Cloning and characterization of a vasopressin V2 receptor and possible link to nephrogenic diabetes insipidus

Stephen J. Lolait*, Anne-Marie O'Carroll*,
O. Wesley McBride†, Monica Konig*, Alain Morel‡
& Michael J. Brownstein*§

* Laboratory of Cell Biology, National Institute of Mental Health,
Building 36, Room 3A-17, Bethesda, Maryland 20892, USA

† Laboratory of Chemistry, National Cancer Institute, NIH, Bethesda,
Maryland 20892, USA

‡ Groupe de Neurobiochimie Cellulaire et Moléculaire,
96 Boulevard Raspail, 75006 Paris, France

THE antidiuretic effect of arginine vasopressin (AVP) is mediated by renal-type (V2) receptors linked to adenyl cyclase¹. We report here the cloning of the rat kidney V2 AVP receptor complementary DNA that encodes a 370-amino-acid protein with a transmembrane topography characteristic of G protein-coupled receptors, and with similarity to the V1a (hepatic) AVP receptor² in its seven membrane-spanning domains. Expression of the cloned cDNA in mammalian cells showed specific ligand binding and activity characteristic of the native V2 AVP receptor. The receptor messenger RNA is detected only in the kidney. The human V2 receptor gene has been localized to the long arm of the X chromosome close to the locus for nephrogenic diabetes insipidus, an X-linked recessive disorder characterized by renal resistance to the antidiuretic action of AVP³⁻⁶.

To clone the rat V2 AVP receptor, two degenerate primers corresponding to putative transmembrane domains II and VI of the rat V1a receptor² were used in a polymerase chain reaction (PCR)⁷ with rat kidney cDNA as a template (see Fig. 1 legend). The sequence of one PCR clone (rkid42) represented a potentially new receptor with high amino-acid similarity to the V1a receptor. Two oligonucleotides complementary to this sequence were used to screen a rat kidney cDNA library. One positive clone (rkid9a) was isolated, but it was not full-length. A full-length receptor cDNA (rkid1.2) was generated using a PCR-based primer extension method (see Fig. 1 legend). Figure 1 shows the nucleotide sequence (1,222 base pairs) and deduced amino-acid sequence of the cDNA. The nucleotide sequence surrounding the initiation codon agrees with the consensus initiation sequence⁸. The longest open reading frame codes for a 370-amino-acid protein (M_r 40,518) that has one potential *N*-linked glycosylation site in the amino terminus and numerous Ser/Thr (12/4) residues in the carboxy terminus that may represent sites for regulatory phosphorylation. Hydropathicity analysis of the translated protein sequence suggests the presence of seven transmembrane domains that are characteristic of G protein-coupled receptors. In these domains, the rkid1.2 cDNA has ~60% sequence identity with the rat V1a receptor, but is not closely related (<25–30% identity in transmembrane domains I–VII) to other peptide receptors^{9–12}.

To identify rkid1.2, the tissue distribution of its transcript was first examined by northern blot analysis. An mRNA of ~2.2 kilobases was found only in the kidney (Fig. 2a). *In situ* hybridization histochemistry revealed high mRNA levels in a pattern corresponding to that of collecting ducts (cortical and medullary), medullary thick ascending limb of loop of Henlé and peri-glomerular tubules (Fig. 2b). This tissue localization is consistent with that of the V2 AVP receptor determined by binding studies and autoradiography^{13,14}.

The pharmacological characteristics of the receptor encoded

TABLE 1 Segregation of V2 AVP receptor gene with human X chromosome

Human chromosome	Gene chromosome				Discordancy (%)
	+/+	+/-	-/+	-/-	
1	19	21	10	16	47
2	14	26	8	18	52
3	19	21	4	22	38
4	30	10	12	14	33
5	19	21	2	24	35
6	30	10	15	11	38
7	15	25	10	16	53
8	22	18	10	16	42
9	18	22	7	19	44
10	13	27	3	23	45
11	22	18	4	22	33
12	23	17	10	16	41
13	15	25	7	19	48
14	25	15	11	15	39
15	22	18	10	16	42
16	22	18	8	18	39
17	25	15	15	11	45
18	18	22	8	18	45
19	16	24	8	18	48
20	26	14	10	16	36
21	31	9	8	18	26
22	21	19	2	24	32
X	40	0	0	26	0

A 224-base-pair human V2 AVP receptor probe (extending from the middle of transmembrane domain II to the beginning of transmembrane domain IV; 5'-GGCTCTGTTCCTCAAGTGTGCTGCCCCAGCTGGCCTGGAAGCGACCGCTCCG-GTGGCCAGATGCCCTGTGTGCGGCCGTGAAGTATCTGCAGATGGTGGCATGTA-TGCCTCTCTACATGATCTGCGCCATGACGCTGGACCGCCACCGTGCATCTGC-CGTCCCATGCTGGCGTACCGCCGTGGAAGTGGGGCTCACTGGAACCGCCCGTGT-TAGTGGC-3'; there is ~90% nucleotide identity between the human and the rat V2 receptor nucleotide sequences in this region) labelled by random priming with [³²P]dCTP, hybridized to a 2.4-kb band in *Bam*HI digests of somatic cell hybrid DNAs. This band was well resolved from 11.7- and 21-kb cross-hybridizing bands in mouse and Chinese hamster DNA digests, respectively. Detection of the human gene is correlated with the presence or absence of each human chromosome in the group of somatic cell hybrids²¹. Discordancy indicates the presence of the gene in the absence of the chromosome (+/-) or absence of the gene despite the presence of the chromosome (-/+), and the sum of these numbers divided by the total number of hybrids examined (×100) represents the per cent discordancy. The human-hamster hybrids consisted of 27 primary clones and 14 sub-clones (29 positive of a total of 41) and the human-mouse hybrids represented 15 primary clones and 10 subclones (11 positive of a total of 25). It was possible to sublocalize the gene to Xq27-qter by examination of hybrids containing only portions of the X chromosome (see text).

by rkid1.2 were studied by expressing the cDNA in eukaryotic cells. Binding of [³H]AVP to membranes prepared from COS-7 cells transiently transfected with the receptor cDNA was saturable with a dissociation constant (K_d) of 0.4 nM and a binding capacity (B_{max}) of ~240 fmol mg⁻¹ protein (Fig. 3a, b). Membranes prepared from control cells did not display specific [³H]AVP-binding activity. Displacement experiments of [³H]AVP binding indicated that AVP, dDAVP (V2 antidiuretic agonist) and dIle²,Ile⁴AVP (V2 antidiuretic antagonist) are equipotent in inhibiting radioligand binding, being about 10- and 100-fold more potent than mVP (V1a pressor antagonist) and oxytocin, respectively (Fig. 3c). The inhibition constants (K_i) for AVP, dDAVP, dIle²,Ile⁴AVP, mVP and oxytocin were calculated to be 1.44, 1.09, 1.1, 17.1 and 93.4 nM, respectively. The rank order of potency obtained agrees well with those reported in binding experiments of the rat kidney V2 AVP receptor¹⁵ and in functional bioassays^{16,17}.

Mouse L cells (containing a cyclic AMP-responsive reporter construct¹⁸) transfected with rkid1.2 cDNA responded to AVP and dDAVP with an increase in cAMP (Fig. 4). The response was completely blocked by the V2 antagonist, dIle²,Ile⁴AVP.

§ To whom correspondence should be addressed.

FIG. 1. Nucleotide sequence for the rklid.2 cDNA and deduced amino-acid sequence of the rat V2 AVP receptor. The nucleotide sequence is numbered at the right hand of each lane. The deduced amino-acid sequence is shown beneath the nucleotide sequence; amino-acid numbering begins with the first methionine of the long open reading frame. An in-frame stop codon in the 5'-untranslated sequence is underlined. A putative *N*-linked glycosylation site at Asn 22 is indicated by an arrow-head. Positions of the putative trans-membrane segments I-VII are indicated by dashed lines above the nucleotide sequence; the termini of each segment are tentatively assigned on the basis of the hydropathy profile and sequence comparison with the rat V1a AVP receptor². The nucleotide sequence has been deposited in the EMBL Data Bank (accession number Z11932).

METHODS. A 1.7×10^6 recombinant rat kidney cDNA library was constructed in the pcD vector of Okayama and Berg. Plasmid DNA (1 μ g) from this library was submitted to 48 cycles of PCR with 1 μ m each of the two degenerate primers (II 5'-CATGCAC(C/T)TCTTTATC-(C/A)GACAC(C/T)TIA(G/A)(C/T)TIGC-3' and VI 5'-T(G/C)(A/T)CCACATCTG-(G/T)AC(A/G)ATGAA(G/A)AA(A/G)GGIGCCCCA-3') based on regions of the rat V1a receptor transmembrane domains II and VI. The timing was 45 s denaturation at 96 °C, 4 min of annealing at 55 °C, and 4 min extension at 72 °C, followed by a final 72 °C extension of 15 min. Four individual bands ranging in size from 420–700 bp were isolated by gel electrophoresis and subcloned into the *HincII* site of M13mp18. A total of 20 clones were sequenced. One clone (rkid42) was especially interesting (see text). Two oligonucleotides (48 bases long) to this sequence were used to screen under stringent conditions (hybridization at $3 \times \text{SSC}$, 60 °C; washing at $1 \times \text{SSC}$, 60 °C) the rat kidney cDNA library by Southern blotting of plasmid DNA from 48 subdivisions of $\sim 50,000$ clones. The one positive cDNA clone isolated (rkid9a; insert size ~ 1.6 kb) was not full-length, stopping at nucle-

When exposed to AVP or dDAVP, cells transfected with the V1a AVP receptor (which is coupled to phosphatidylinositol metabolism¹⁹) cDNA showed no elevation of cAMP. These expression data thus confirm that the cDNA that we have cloned encodes a functional V2 AVP receptor protein.

A defect in the V2 receptor has been implicated in the disorder nephrogenic diabetes insipidus⁴. In humans, this rare X-linked recessive condition is characterized by the kidney's inability to

	tgtctgtacaggtgtgt	L aggtcatcatcaaccacccatc	-1

tttttaccocctagctctccagcaacagcagccaggaggaactactgtagtcagagaccgcgtgtagtcgcggtcgtaa			120
PHSerProSerProSerSerProSerAsnSerSerGlnGluGluLeuLeuAspAspArgAspProLeuLeuValArgAlaGlu			40
	▲		

agacatggcctagctgctgggggcctaatcagcggcgccggcgctggacgctgggcacccatgcattctctcatcagctac			240
SerAsnGlyLeuValLeuGlyAlaLeuIleArgArgGlyArgArgGlyArgTrpAlaProMetHisValPheIleSerHis			80

gctaccccaagctggctgggtagcgaactgacgcgcttccatggccctgatgcctgtgtcgccgctgaagtacctgcagatg			360
LeuProGlnLeuAlaTrpAspAlaThrAspArgPheHisGlyProAspAlaLeuCysArgAlaValIysTyrLeuGlnMet			120

gacactagacgcgccaactgctgccatctgcgcacctatgctagcataccgccaatggaggtggggctcgctggaacaggccagctg			480
ThrLeuAspArgHisArgAlaIleCysArgProMetLeuAlaTyrArgHisGlyGlyAlaArgTrpAsnArgProVal			160

cagctcttcatcttctgctcagcgtgatgtgggaatggcagctggggctgttgattgctgggcccgatttgcagaacccatgg			600
GlnLeuPheIlePheAlaGlnArgAspValGlyAsnGlySerGlyValPheAspCysTrpAlaArgPheAlaGluProTrp			200

ttttgtggcacctgcacctagccattgctgcctgtcaggttcttattcttccggagatcacaccagctctgtgtgccaggcca			720
PheValAlaProAlaLeuGlyIleAlaAlaCysGlnValLeuIlePheArgGluIleHisAlaSerLeuValProGlyPro			240

agccccagcaggggagcacatgtatcagcagccatggccaagccgtgagatgacactgggtgattgtgattgtctacgtg			840
SerProSerGluGlyAlaHisValSerAlaAlaMetAlaIysThrValArgMetThrLeuValIleValIleValTyrVal			280

gcgtgggactcgggaagctcctctggaaaagccccctttgtgtgctcatgctgcgtgcgtgaccttaacagctgtaccaac			960
AlaTrpAspProGluAlaProLeuGluArgProProPheValLeuLeuMetLeuLeuAlaSerLeuAsnSerCysThrAsn			320

gagttgcgtagcctgcttctgctgtcctcagaggcacaccacacacagcctgggtcctcaagatgaatcctgtgccacagcc			1080
GluLeuArgSerLeuLeuGlyCysAlaGlnArgHisThrThrHisSerLeuGlyProGlnAspGluSerCysAlaThrAla			360

actgggctcttcttactcaaaacccctctgtagctcaactgactatctagggttggccctga			1222
			370

tide number 34 of the coding sequence shown above. Additional 5'-cDNA sequence was obtained using PCR on 1 µg plasmid DNA from the rat kidney cDNA library and the RACE (rapid amplification of cDNA ends) technique²³ with nested primers; the two upstream primers were vector-specific. The final full-length cDNA of 1,222 bp (rkid1.2) was generated from the cDNA library by PCR, using sequence generated by the RACE procedure as the 5'(upstream) primer (the first 24 bp of the cDNA: 5'-TGCTGTGACAGGTGTGTT-AGGTCA-3') and sequence 3' from the stop codon derived from the short clone (rkid9a) as the 3' (downstream) primer (last 24 bp of the cDNA: 5'-TCAGGCCAACCCCTAGATAGTCAG-3'). The cDNA sequence was obtained by sequencing both the partial rat cDNA clone (rkid9a) and the full-length PCR cDNA (rkid1.2). Both strands were sequenced by the Sanger dideoxynucleotide chain termination method using Sequenase (USB).

concentrate urine resulting in polyuria, polydipsia and urine of low specific gravity. The nephrogenic diabetes insipidus locus has been assigned to band Xq28 near the long-arm telomere of the X chromosome^{3,6}. More recently, a close correlation between the presence of the nephrogenic diabetes insipidus locus and V2 receptor expression (measured biochemically) has been documented²⁰. A partial-length human V2 receptor cDNA was obtained by screening a human kidney cDNA library with a

FIG. 2 *a*, Northern blot analysis of the tissue distribution of *rkid1.2* mRNA. The ~2.2 kb receptor mRNA is indicated by a large arrowhead. Each lane contained 5 μ g of poly(A)⁺ RNA from rat uterus (Ut), seminal vesicles (Sv), cerebellum (Cb), pituitary (Pit), liver (Li) and kidney (Kl). Numbers on the right represent RNA size markers (BRL). Only the kidney expresses the receptor mRNA after an overnight (top) or five-day (bottom) exposure. *b*, Localization of *rkid1.2* mRNA in rat kidney by *in situ* hybridization histochemistry. The cortical (cx) and medullary (m) regions expressing receptor mRNA appear white in the photomicrograph. Exposure against hyperfilm $\beta_{\max}$ (Amersham) was for 3 weeks.

METHODS. Northern blot analysis² and *in situ* hybridization histochemistry^{24,25} were as previously described. A ³²P-labelled 48-bp oligonucleotide (complementary to bases 415–462 of rkid1.2) was used as a probe for northern analysis; the probe was ³⁵S-labelled for *in situ* hybridization

A black and white micrograph showing a dense, granular structure, likely a biological specimen. The image is labeled with 'CX' at the top center and 'm' in the middle right. The structure appears to be a cross-section of a tissue or a cluster of cells, with a central region labeled 'm' and a surrounding region labeled 'CX'.

histochemistry. Hybridization of serial sections with a ³⁵S-labelled sense probe (bases 415–462 of rkid1.2) resulted in uniform background labelling (data not shown). The experiments were done 2–4 times with similar results.

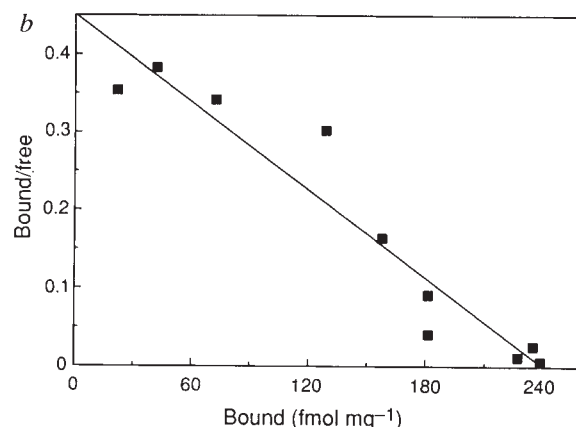
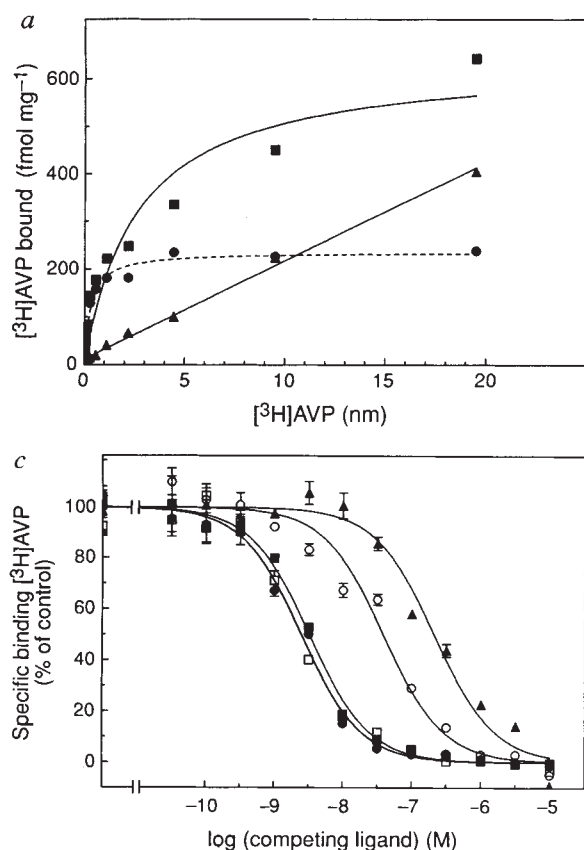


FIG. 3 Binding of $[^3\text{H}]\text{AVP}$ to membranes from COS-7 cells transfected with clone rkd1.2. **a**, Saturation isotherm of total (■), specific (●) and nonspecific (▲) binding of $[^3\text{H}]\text{AVP}$ to membranes from COS-7 cells expressing the rkd1.2 cDNA. Nonspecific binding was determined in the presence of 1 μM unlabelled AVP. **b**, Scatchard plot of the specific binding data in **a**. The estimated maximal binding (B_{max}) and K_d were 240 fmol mg^{-1} protein and 0.4 nM, respectively. **c**, Pharmacological specificity of $[^3\text{H}]\text{AVP}$ binding to COS-7 cell membranes. Representative curves are shown for the concentration-dependent inhibition of $[^3\text{H}]\text{AVP}$ binding by AVP (■), dDAVP (1-deamino, 8-D-arginine vasopressin; □), dIle²,Ile⁴AVP ((d(CH₂)₅-D-Ile²,Ile⁴,Arg⁸)-vasopressin) (●), mVP (1-mercapto- β , β -cyclopentamethylene propionic acid (O-methyl-tyrosine)8-arginine) vasopressin; (d(CH₂)₅Tyr(Me)AVP) (○) and oxytocin (▲). All peptides were purchased from Peninsula. Competition experiments used 0.5 nM $[^3\text{H}]\text{AVP}$.

METHODS. The full-length rkd1.2 cDNA was blunt-ended and kinased with T4 polymerase and T4 polynucleotide kinase, respectively (NEB), and cloned into the blunt-ended *Xho*I site of the plasmid expression vector, pcDM7/*Bam*HI⁻ (courtesy of B. Seed). The appropriate 5'-3' orientation of the clone was verified by restriction mapping. The construct was used transiently to transfect COS-7 cells by the calcium phosphate precipitation technique²⁶. $[^3\text{H}]\text{AVP}$ binding and data analysis were as described previously^{27,28}. Results shown are means ($\pm$ s.e.m.) in **c** of triplicate determinations. The saturation and competition experiments were done twice with similar results.

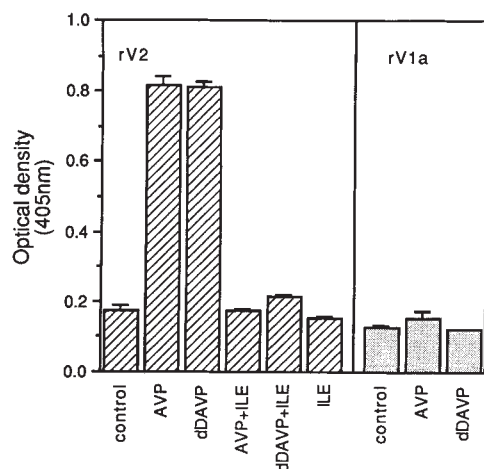


FIG. 4 V2 AVP receptor stimulation of cAMP production in mouse L cells containing a cAMP responsive reporter construct. In this assay¹⁸, cells transfected with a G_s -coupled receptor cDNA activate adenylate cyclase, increase intracellular cAMP, and induce β -galactosidase activity detected by staining cells with a chromogenic substrate. Colour development is measured in a 96-well plate reader at 405 nm. The cAMP response of the rkd1.2 construct (rV2) on addition of 10^{-8} M AVP or dDAVP is completely blocked by the V2 antagonist (d(CH₂)₅-D-Ile²,Ile⁴,Arg⁸)-vasopressin (ILE; 10^{-6} M). There is no response on application of the antagonist alone; similarly, there is no AVP- or dDAVP-induced cAMP response of the rat V1a receptor (rV1a). Control is the optical density reading with no drug addition.

³²P-labelled *Nsi*I-*Kpn*I (~900 base pairs) fragment of the rkd9a clone (data not shown). Southern blots of the DNAs from 10 unrelated individuals were examined with a ³²P-labelled human V2 receptor cDNA probe (see Table 1 legend) after separate digestions with 11 different restriction enzymes (*Eco*RI, *Hin*dIII, *Bam*HI, *Xba*I, *Sac*I, *Taq*I, *Pvu*II, *Pst*I, *Bgl*II, *Eco*RV and *Kpn*I), and no restriction fragment length polymorphism was detected. Only a single band was found in digests with nine restriction endonucleases, indicating that the V2 probe detects a single copy gene in the human genome.

We have localized the V2 AVP receptor gene to the human X chromosome by Southern analysis of a group of human-rodent somatic cell hybrid DNAs digested with *Bam*HI (Table 1). The results were confirmed by analysis of *Eco*RI and *Hin*dIII digests of these same hybrid cell DNAs. This localization was strongly supported by the detection of the human gene in one human-mouse hybrid containing only the human X chromosome and a portion of 2p and in a human-hamster hybrid

containing only human chromosome 22, 6pter-q13 and X. Regional localization was determined by analysis of hybrids containing translocations and breaks involving the X chromosome. One series of human-hamster hybrids²¹ was isolated after fusion of human fibroblasts (GM0073) containing a reciprocal X:14 (q13;q32) chromosome translocation. The V2 receptor cDNA probe hybridized to 10 independent hybrids retaining only the Xq13-qter translocation chromosome but not to three independent hybrids retaining the Xpter-q13 translocation chromosome in the absence of the normal X or the reciprocal translocation chromosome. The DNAs isolated from a series of mouse cell lines containing smaller regions of the human X chromosome after chromosome-mediated gene transfer²² were also examined. Six lines containing only small portions of human X chromosome including the hypoxanthine phosphoribosyl-transferase (*hprt*) gene (Xq26) did not contain the V2 receptor gene, whereas another line (A9S2) containing the distal third of the X chromosome including *hprt* and the glucose-6-phos-

phate dehydrogenase gene (Xq28) retained the V2 receptor gene. These results suggest that the V2 AVP receptor gene is located at Xq27-q28, although a region just proximal to Xq26 cannot be rigorously excluded. Definite linkage between the V2 AVP receptor and nephrogenic diabetes insipidus will require the characterization of a mutant V2 receptor gene from an affected individual. □

Received 18 February; accepted 31 March 1992.

- Jard, S. *Curr. Top. Membranes Transp.* **18**, 255-285 (1983).
- Morel, A., O'Carroll, A.-M., Brownstein, M. J. & Lolait, S. J. *Nature* **356**, 523-526 (1992).
- Bode, H. H. & Crawford, J. D. *New Engl. J. Med.* **280**, 750-754 (1969).
- Bichet, D. G. *et al. Kidney Int.* **36**, 859-866 (1989).
- Knoers, N. *et al. Hum. Genet.* **80**, 31-38 (1988).
- Knoers, N. *et al. Nephron* **50**, 187-190 (1988).
- Saiki, R. K. *et al. Science* **239**, 487-491 (1988).
- Kozak, M. *Nucleic Acids Res.* **12**, 857-872 (1984).
- Arai, H., Hori, S., Aramori, I., Ohkubo, H. & Nakanishi, S. *Nature* **348**, 730-732 (1990).
- Lin, H. Y. *et al. Science* **254**, 1022-1024 (1991).
- Juppner, H. *et al. Science* **254**, 1024-1026 (1991).
- Yamada, Y. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 251-255 (1992).
- Jard, S., Barberis, C., Audigier, S. & Tribollet, E. *Prog. Brain Res.* **72**, 173-187 (1987).
- Phillips, P. A. *et al. Endocrinology* **126**, 1478-1484 (1990).
- Dorsa, D. M., Majumdar, L. A., Petracca, F. M., Baskin, D. G. & Cornett, L. E. *Peptides* **4**, 699-706 (1983).
- Manning, M. *et al. J. med. Chem.* **27**, 423-429 (1984).
- Kruszynski, M. *et al. J. med. Chem.* **23**, 364-368 (1980).
- König, M., Mahan, L. C., Marsh, J. W., Fink, J. S. & Brownstein, M. J. *Molec. cell. Neurosci.* **2**, 331-337 (1991).
- Kirk, C. J., Verrinder, T. R. & Hems, D. A. *FEBS Lett.* **83**, 267-271 (1977).
- Jans, D. A., van Oost, B. A., Ropers, H. H. & Fahrenholz, F. *J. biol. Chem.* **265**, 15379-15382 (1990).
- McBride, O. W. *et al. Nucleic Acids Res.* **10**, 8155-8170 (1982).
- Olsen, A. S., McBride, O. W. & Moore, D. E. *Molec. cell. Biol.* **1**, 439-448 (1981).
- Frohman, M. A., Dush, M. K. & Martin, G. R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8998-9002 (1988).
- Young, W. S. III, Bonner, T. I. & Brann, M. R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9827-9831 (1986).
- Young, W. S. III, Warden, M. & Mezey, E. *Neuroendocrinology* **46**, 439-444 (1987).
- Chen, C. & Okayama, H. *Molec. cell. Biol.* **7**, 2745-2752 (1987).
- Holmes, M. C., Antoni, F. A. & Szentendrei, T. *Neuroendocrinology* **117**, 421-423 (1985).
- Antoni, F. A. *Neuropeptides* **4**, 413-420 (1984).

ACKNOWLEDGEMENTS. We thank L. Mahan for assistance in analysing the binding data and W. S. Young and A. Spiegel for discussions. A.-M. O'C. was supported in part by a grant from Ciba Geigy, and S.J.L. was supported by grants from the National Alliance for Research on Schizophrenia and Depression and the Stanley Foundation.

Changing role of *even-skipped* during the evolution of insect pattern formation

Nipam H. Patel^{*†}, Eldon E. Ball^{*‡}
& Corey S. Goodman^{*}

^{*} Howard Hughes Medical Institute, Department of Molecular and Cell Biology, 519 LSA, University of California, Berkeley, California 94720, USA

[†] Carnegie Institution, 115 West University Parkway, Baltimore, Maryland 21210, USA

[‡] Molecular Neurobiology Group, Research School of Biological Sciences, PO Box 475, Canberra, Australian Capital Territory 2601, Australia

THE development of *Drosophila* is typical of the so-called long germband mode of insect development, in which the pattern of segments is established by the end of the blastoderm stage^{1,2}. Short germband insects, such as the grasshopper *Schistocerca americana*, by contrast, generate all or most of their metameric pattern after the blastoderm stage by the sequential addition of segments during caudal elongation³. This difference is discernible at the molecular level in the pattern of initiation of the segment polarity gene *engrailed*⁴, and the homeotic gene *abdominal-A* (ref. 5). For example, in both types of insects, *engrailed* is expressed by the highly conserved germband stage^{4,6} in a pattern of regularly spaced stripes, one stripe per segment⁷⁻⁹. In *Drosophila*, the complete pattern is visible by the end of the blastoderm stage, although *engrailed* appears initially in alternate segments in a pair-rule pattern^{9,10} that reflects its known control by pair-rule genes such as *even-skipped*¹¹⁻¹⁵. In contrast, in the grasshopper, the *engrailed* stripes appear one at a time after the blastoderm stage as the embryo elongates⁴. To address the molecular basis for this

difference, we have cloned the grasshopper homologue of the *Drosophila* pair-rule gene *even-skipped* and show that it does not serve a pair-rule function in early development, although it does have a similar function in both insects during neurogenesis later in development.

Figure 1 shows a comparison of the proteins encoded by the *Drosophila even-skipped* gene and the putative grasshopper homologue. A high degree of amino-acid identity exists within the homeodomain (56 out of 60 amino acids), in the adjacent C-terminal region (region A in Fig. 1; 17 out of 24 amino acids), and in a short stretch at the C-terminal end of the protein (region B in Fig. 1; 7 out of 10 amino acids). Comparisons of the grasshopper sequence with other homeodomain-containing proteins of *Drosophila* reveal that no other homeodomain has more than 60% amino-acid identity and none has significant sequence similarity outside of the homeodomain. Thus the sequence comparisons indicate that the grasshopper gene described here is a homologue of the *Drosophila even-skipped* gene.

We used antibodies raised against grasshopper *even-skipped* to investigate the protein distribution during development. Midway through *Drosophila* embryogenesis, *even-skipped* is expressed in a segmentally repeated subset of identified neurons, an ectodermal ring at the anal pad, and in segmentally reiterated mesodermal cells at the dorsal edge of the embryo^{13,14,16} (Fig. 2a). At an equivalent stage of development, grasshopper *even-skipped* is expressed in an almost identical pattern (Fig. 2b). The neural expression of *even-skipped* is known to play an important part in the differentiation of particular identified neurons in *Drosophila*¹⁶, homologues of which are readily identifiable in grasshopper¹⁷. A careful examination suggests that *even-skipped* is expressed in homologous neurons in the nervous systems of both insects; for example, on the dorsal surface of the central nervous system (CNS), *even-skipped* is expressed by only the aCC, pCC and RP2 neurons (Fig. 2c and d). The similarities in expression pattern, particularly in the nervous system, are further evidence that we have indeed isolated the grasshopper homologue of *Drosophila even-skipped*.

To analyse the potential pair-rule function of grasshopper *even-skipped*, we examined embryos at a much earlier stage of development. During the blastoderm stage, *Drosophila even-skipped* expression is resolved into a series of seven stripes^{13,14}. Shortly thereafter, *engrailed* expression is initiated in a pattern of fourteen stripes, with the anterior margin of each *even-skipped* stripe coinciding with an odd-numbered *engrailed* stripe¹². At the onset of gastrulation, *even-skipped* stripes have sharply defined anterior borders and *engrailed* stripes show a pair-rule pattern of intensity, reflecting their order of initiation (Fig. 3a and d). Detailed genetic analysis has indicated that the pair-rule pattern of *Drosophila even-skipped* expression is required to establish the proper pattern of *engrailed* expression^{11,13,15}.

To determine the potential for *even-skipped* regulation of *engrailed* in grasshopper, we examined the expression of *even-skipped* at stages before (12% to 16% of development) and during (17% to 30%) the appearance of *engrailed* stripes⁴. Figure 3b and c shows an embryo in which the first two *engrailed* stripes (T1 and T2) have just started to appear. At this stage, grasshopper *even-skipped* (Fig. 3e and f) does not show any pair-rule pattern of expression and does not overlap with *engrailed* expression. At no time in development do we observe an expression pattern that would suggest that grasshopper *even-skipped* is involved in the initiation of *engrailed* striped expression. In fact, with the exception of a few neurons, we have not observed grasshopper *even-skipped* and *engrailed* localization within the same nucleus.

If pair-rule patterning were involved in short germband development, one might expect to find grasshopper homologues of at least two of the three *Drosophila* primary pair-rule genes (*hairy*, *runt* and *even-skipped*) expressed in a pair-rule fashion before the appearance of *engrailed* stripes. Grasshopper

even-skipped, however, does not display a pair-rule pattern of expression and therefore does not appear to play a part in setting up the pattern of *engrailed* expression during grasshopper development. These observations strengthen the argument that pair-rule patterning is not involved in the segmentation of short germband embryos, such as grasshopper, which are representative of the primitive mode of insect development. One caveat to this analysis would be the existence of a second *even-skipped* gene in grasshopper that does play a part in pair-rule patterning, but we have not found any indication of such a second gene from our polymerase chain reaction experiments, nor from the analysis of genomic Southern blots probed with the existing grasshopper *even-skipped* gene.

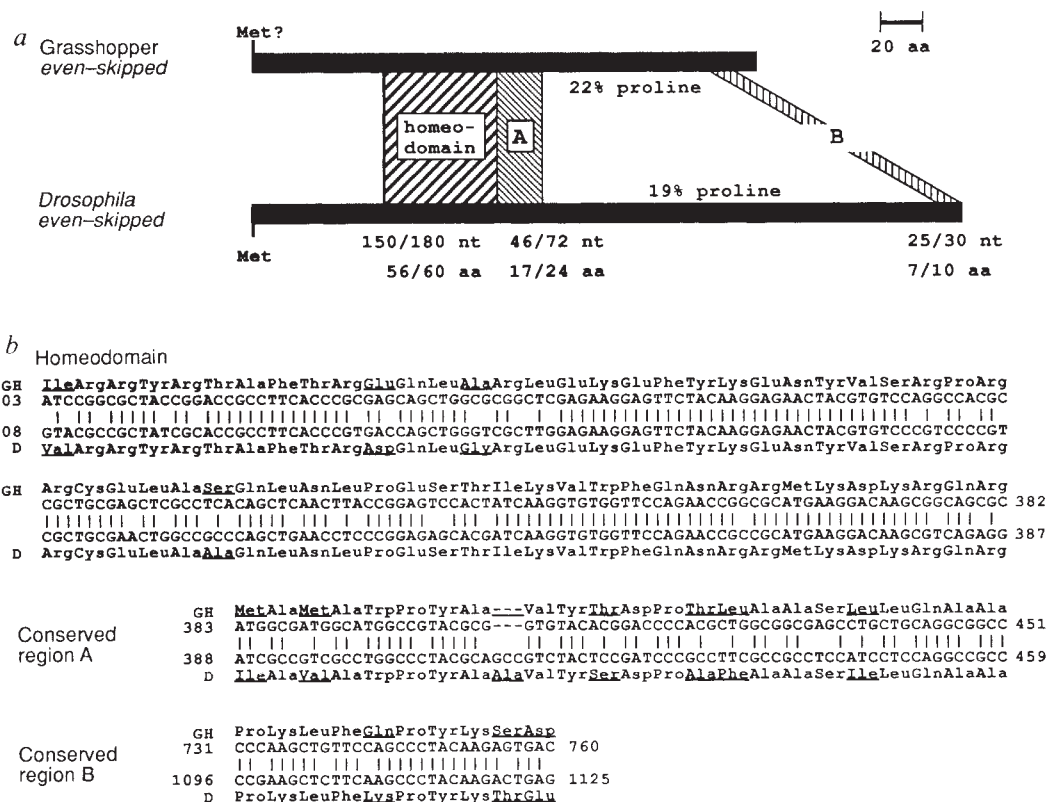
Despite the lack of pair-rule expression, grasshopper *even-skipped* is expressed in the early embryo. At the onset of gastrulation, expression is observed in the posterior region of the embryo and is predominantly mesodermal, with low levels in the overlying ectoderm. At no point does this expression pattern resolve into a pattern of stripes. As the embryo begins to elongate, expression is maintained in the most posterior region of mesoderm (Fig. 3e), with the highest levels of expression in

mesodermal cells in medial regions of the gastral groove. At about 25% of development, this mesodermal staining disappears. Thus, early expression in grasshopper seems to be limited to posterior mesoderm during the time that the embryo is gastrulating—an expression pattern unlike anything reported for *Drosophila even-skipped*^{13,14}.

This expression in posterior mesoderm, however, is reminiscent of the early expression patterns of *Xhox-3* and *Eux-1*, *Xenopus* and mouse *even-skipped* homologues, respectively¹⁸⁻²⁰. For these vertebrate genes, expression is predominantly in the posterior mesoderm; a variety of manipulations in *Xenopus* suggest that they may be involved in early axial patterning^{21,22}. Later in development, these vertebrate *even-skipped* homologues are expressed within the central nervous system^{18,19}. Interestingly, the vertebrate *even-skipped* homologues appear to be associated with Hox gene clusters. For example, the human *even-skipped* homologues, *EVX-1* and *EVX-2*, are located at the extreme 5' ends of the clusters on chromosomes 7 and 2, respectively^{23,24}. In *Drosophila*, however, *even-skipped* is located on a different chromosome from the one containing the Antennapedia and Bithorax complexes. We do not yet know if the

FIG. 1 Alignment of *Drosophila* and grasshopper *even-skipped* sequences. a, Schematic representations of the grasshopper and *Drosophila even-skipped* proteins^{15,16} indicating the three regions of amino-acid sequence similarity: the homeodomain, region A (extended homeodomain) and region B. Below each region, the ratios of nucleotide (nt) and amino-acid (aa) identities are given. Between regions A and B, the predicted sequence of both proteins indicates a high proportion of proline residues. b, Nucleotide and amino-acid alignments of the three regions of similarity. Amino-acid mismatches are underlined and a single amino-acid gap has been inserted in grasshopper region A (dashed lines) for the purposes of sequence alignment. For *Drosophila even-skipped*, nucleotide numbering (shown at the beginning and end of each domain) begins with the adenine of the start methionine codon of the protein^{13,14}. For the grasshopper sequence, numbering begins with the first nucleotide of the cDNA sequence after the *EcoRI* cloning site. There is an in-frame methionine starting at the fifth base pair, but the surrounding nucleotides match poorly with the consensus seen around start methionine codons²⁵; thus we believe that the actual start methionine may not be included in our cDNA clone.

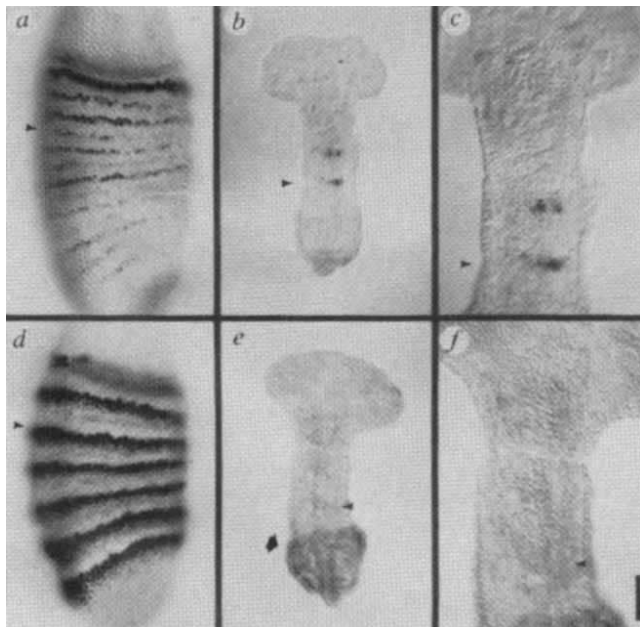
METHODS. mRNA was isolated from grasshopper embryos ranging from 28% to 35% of embryonic development. First strand cDNA, synthesized using oligo(dT) and M-MLV reverse transcriptase, was used as a template for PCR reactions²⁶. Based on the comparison of the homeobox sequences of *Drosophila eve* and mouse *Eux-1* with other homeoboxes, four different primers were made to amplify *even-skipped* class homeoboxes specifically by polymerase chain reaction (PCR): (1) 5'-GGCAAGCTTGGARAARGARTYTYA-3'; (2) 5'-CGCCTCGAGAARGARTTYTAYARGGARAAYT-3'; (3) 5'-CATCGCGGRTTYTGRAACANACYTTXAT-3'; and (4) 5'-TTTGATCCYTGNCGYTTRTCYTTTCAT-3' (where N is A,C,G or T; R is A or G; Y is C or T; and X is A,G or T). Primers were used at a 1 µM final concentration in a total



volume of 25 µl. Primers 1 and 4 were used for 35 cycles (95 °C for 1 min, 55 °C for 1 min, 72 °C for 45 s), and 2.5 µl of this product was then used as the template for another 35 cycles of amplification using primers 2 and 3. The final product was purified after electrophoresis in a 1.9% agarose gel, kinased, subcloned into M13, and sequenced. Twelve M13 clones were sequenced, and all contained the identical amplification product (corresponding to nucleotides 276–335 of the grasshopper *even-skipped* cDNA sequence shown above). This product was also used to screen a total of 5 × 10⁶ plaques from grasshopper λgt11 and λgt10 cDNA libraries²⁷. No positives were found from three different pools of a λgt11 library, but a screen of 1 × 10⁶ plaques from the λgt10 library yielded three identical cDNA clones. The grasshopper *even-skipped* cDNA was sonicated, and fragments were then subcloned into M13 and sequenced. The total length of the cDNA is 2,140 base pairs, of which 1,341 base pairs are 3' untranslated sequence. The complete grasshopper *even-skipped* cDNA sequence is available through the EMBL, GenBank and DDBJ Databases (accession number Z11845).

FIG. 2 Comparison of *even-skipped* expression in *Drosophila* and grasshopper at mid-embryogenesis. *a* and *c*, Expression of *even-skipped*, as visualized with monoclonal antibody 3C10, in a *Drosophila* embryo at 12 h of development; *b* and *d*, *even-skipped* expression in a 45% grasshopper embryo as detected with an affinity-purified rat antiserum. In both *Drosophila* (*a*) and grasshopper (*b*) embryos, *even-skipped* protein accumulates in a subset of neurons (large arrowhead), in dorsal mesoderm (open arrow) and in an ectodermal ring at the anal pad (small arrowhead in *a*; not shown in *b*). The *Drosophila* embryo appears to have more neural staining, but this is simply because the greater thickness of the grasshopper embryo prevents visualizing all stained neurons in a single focal plane. Homologous neurons express *even-skipped* in both insects. For example, *c* and *d* show the dorsal-most neurons that express *even-skipped* in *Drosophila* (*c*) and grasshopper (*d*). In both, *even-skipped* protein is detected in RP2 (arrowhead), aCC (wide arrow), and pCC (thin arrow), but not in any of the other well-characterized dorsal neurons. Scale bar, 70 μ m (*a* and *b*), 8 μ m (*c*), 10 μ m (*d*). Anterior is up in all panels.

METHODS. To produce antibodies against grasshopper *even-skipped*, the cDNA was subcloned into the *Eco*RI site of Bluescript KS⁺, and then a *Sac*I to *Xba*I fragment (*Sac*I site at nucleotide position 304 of grasshopper *even-skipped*, *Xba*I site provided by the Bluescript polylinker) was subcloned into both pRIT31 (ref. 28) and pATH11 (ref. 29) cut with *Sac*I and *Xba*I. These were used to generate protein A and *trpE* fusion proteins containing the C-terminal two-thirds of grasshopper *even-skipped*. The entire *Eco*RI fragment was also subcloned into the *Eco*RI site of pGEX-3X (Pharmacia) to generate a glutathione *S*-transferase fusion protein containing the entire protein encoded by the cDNA. The protein A fusion (50 μ g) was injected into rats at two-week intervals. After four injections, the resulting antiserum was affinity-purified on a column of the *trpE* fusion protein coupled to Affigel 10/15 (Bio-Rad) and eluted with glycine-HCl, pH 2.3. Fractions were tested on western blot strips of the glutathione *S*-transferase fusion. For the production of a monoclonal antibody against *Drosophila even-skipped*, a plasmid producing *Drosophila even-skipped* under the control of the T7 promoter was provided by T. Hoey and M. Levine. The bacterially produced *Drosophila even-skipped* protein was used to immunize mice and generate monoclonal antibody 3C10. All immunohistochemistry followed published methods³⁰. For both 3C10 and the rat antiserum, staining was improved by limiting the fixation time to 10 min.



grasshopper *even-skipped* gene is located near any of the homeotic gene homologues.

Our results provide some initial answers to long-standing questions concerning the evolution of insect segmentation. We suggest that *even-skipped* had a role in neurogenesis and/or axial patterning in the common ancestor to vertebrates and arthropods. The extremely similar expression patterns in identified neurons, dorsal mesoderm and the ring of tissue at the anal pad, suggest a highly conserved role in the germband

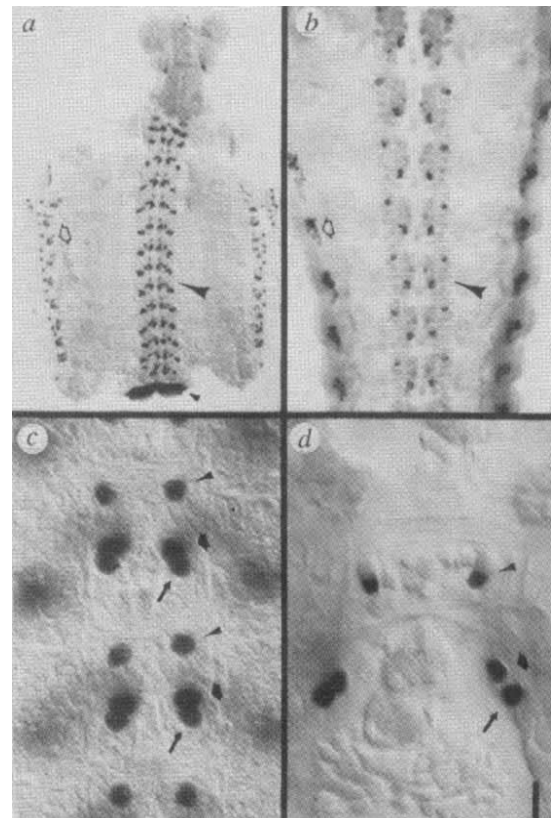


FIG. 3 Expression of *even-skipped* during *Drosophila* and grasshopper segmentation. *Drosophila engrailed* (*a*) and *even-skipped* (*d*) protein accumulation at the end of gastrulation shows that *even-skipped* stripes (*d*) are sharply defined at the time that *engrailed* stripes (*a*) are forming. Arrowhead marks the position of the posterior portion of the second thoracic (T2) segment (*engrailed* stripe 5, anterior part of *even-skipped* stripe 3). The anterior margin of each *even-skipped* stripe corresponds to each of the odd-numbered *engrailed* stripes^{1,2}. Note the pair-rule pattern of stripe intensity in which the odd-numbered *engrailed* stripes are weaker than the even-numbered ones at this stage. At 17% of grasshopper development, *engrailed* stripes (*b*, and at higher magnification in *c*) have just appeared in T1 and T2 (arrowhead marks T2). The next stripes to appear will be further anterior in S1 and S3 (the mandibular and labial segments). The stripes of S1 and S3 represent the only pair of grasshopper *engrailed* stripes to appear in any pattern reminiscent of the pair-rule pattern of initiation seen for *Drosophila engrailed*⁴. In grasshopper embryos of the same stage stained for *even-skipped* expression (*e* and *f*), no pair-rule stripe pattern is visible. The position of T2, which can be determined even in unstained preparations by bulges in the mesoderm which are visible shortly after *engrailed* stripes appear, is marked by an arrowhead. The higher magnification view in *f* shows that no stripe of *even-skipped* expression is evident in T2, where *engrailed* expression has started, or more anteriorly in S1 and S3, where *engrailed* stripes will next appear. There is, however, a domain of grasshopper *even-skipped* expression in the more posterior mesoderm (arrow in *e*). At no earlier or later stage do we observe *even-skipped* expression in a pattern of ectodermal stripes that would be consistent with a role in the regulation of *engrailed* expression. Scale bar represents 200 μ m (*b* and *e*), 100 μ m (*c* and *f*), and 75 μ m (*a* and *d*). Anterior is up in all panels.

stage of all insects. During the course of insect evolution, insects using a long germband mode of development arose and *even-skipped* acquired an additional function, that of pair-rule patterning. □

Received 14 February; accepted 8 April 1992.

1. Ingham, P. *Nature* **335**, 25–34 (1988).
2. Akam, M. *Development* **101**, 1–22 (1987).
3. Sander, K. *Adv. Insect Physiol.* **12**, 125–238 (1976).

4. Patel, N. H., Kornberg, T. B. & Goodman, C. S. *Development* **107**, 201–212 (1989).
5. Tear, G., Akam, M. & Martinez-Arias, A. *Development* **110**, 915–925 (1990).
6. Sander, K. *Development* **104**, (suppl.) 112–121 (1988).
7. Kornberg, T., Siden, I., O'Farrell, P. H. & Simon, M. *Cell* **40**, 45–53 (1985).
8. Fjose, A., McGinnis, W. & Gehring, W. J. *Nature* **313**, 284–289 (1985).
9. DiNardo, S., Kuner, J., Theis, J. & O'Farrell, P. H. *Cell* **43**, 59–69 (1985).
10. Weir, M. P. & Kornberg, T. *Nature* **318**, 433–439 (1985).
11. DiNardo, S. & O'Farrell, P. H. *Genes Dev.* **1**, 1212–1225 (1987).
12. Lawrence, P. A., Johnston, P., Macdonald, P. & Struhl, G. *Nature* **328**, 440–442 (1987).
13. Macdonald, P., Ingham, P. & Struhl, G. *Cell* **47**, 721–734 (1986).
14. Frasch, M., Hoey, T., Rushlow, C., Doyle, H. & Levine, M. *EMBO J.* **6**, 749–759 (1987).
15. Frasch, M., Warrior, R., Tugwood, J. & Levine, M. *Genes Dev.* **2**, 1824–1838 (1988).
16. Doe, C. Q., Smouse, D. & Goodman, C. S. *Nature* **333**, 376–378 (1988).
17. Thomas, J. B., Bastiani, M. J., Bate, C. M. & Goodman, C. S. *Nature* **310**, 203–207 (1984).
18. Ruiz i Altaba, A. & Melton, D. A. *Development* **106**, 173–183 (1989).
19. Bastiani, H. & Gruss, P. *EMBO J.* **9**, 1839–1852 (1990).
20. Dush, M. K. & Martin, G. M. *Dev. Biol.* (in the press).
21. Ruiz i Altaba, A. & Melton, D. A. *Cell* **57**, 317–326 (1989).
22. Ruiz i Altaba, A. & Melton, D. A. *Trends Genet.* **6**, 57–64 (1990).
23. D'Esposito, M. et al. *Genomics* **10**, 43–50 (1991).
24. Boncinelli, E., Simeone, A., Acampora, D. & Mavilio, F. *Trends Genet.* **7**, 329–334 (1991).
25. Cavener, D. R. *Nucleic Acids Res.* **15**, 1353–1361 (1989).
26. Grenningloh, G., Rehm, E. J. & Goodman, C. S. *Cell* **67**, 45–57 (1991).
27. Snow, P. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5291–5295 (1988).
28. Nilsson, B. & Abrahmsen, L. *Meth. Enzym.* **185**, 144–161 (1990).
29. Koerner, T. J., Hill, J. E., Meyers, A. M. & Tzagoloff, A. *Meth. Enzym.* **194**, 477–490 (1991).
30. Patel, N. H. et al. *Cell* **58**, 955–968 (1989).

ACKNOWLEDGEMENTS. We thank M. Dush and G. Martin for sharing results before publication and for discussion on PCR primers; K. Zinn for the grasshopper libraries; G. Grenningloh, J. Rehm and M. Hortsch for advice; G. Tear and Y. Hiromi for discussion; R. Chasan for comments on the manuscript; T. Hoey and M. Levine for the *Drosophila eve* expression construct; and B. Blankenmeier and K. Jepson-Innes for technical assistance. This work was supported by an NIH grant to C.S.G. who is an investigator with the Howard Hughes Medical Institute.

Location of MHC-encoded transporters in the endoplasmic reticulum and *cis*-Golgi

Monique J. Kleijmeer*, Adrian Kelly†, Hans J. Geuze*, Jan W. Slot*, Alain Townsend‡ & John Trowsdale§†

* Department of Cell Biology, School of Medicine, University of Utrecht, The Netherlands

† Human Immunogenetics Laboratory, Imperial Cancer Research Fund, Lincoln's Inn Fields, Holborn, London WC2A 3PX, UK

‡ Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

IMMUNE recognition of intracellular proteins is mediated by major histocompatibility complex (MHC) class I molecules that present short peptides to cytotoxic T cells^{1–4}. Evidence suggests that peptides arise by cleavage of proteins in the cytoplasm and are transported by a signal-independent mechanism into a pre-Golgi region of the cell, where they take part in the assembly of class I heavy chains with β_2 -microglobulin (reviewed in refs 5–7). Analysis of cells that have defects in class I molecule assembly and antigen presentation^{8–14} has shown that this phenotype can result from mutations in either of the two ABC transporter genes located in the class II region of the MHC^{15–22}. This suggested that the protein complex encoded by these two genes^{20,22} transports peptides from the cytosol into the endoplasmic reticulum. Here we report additional evidence by showing that the transporter complex is located in the endoplasmic reticulum membrane and is probably oriented with its ATP-binding domains in the cytosol.

To locate the MHC-encoded transporter proteins, an anti-serum (AK1.7) was raised against the carboxy-terminal peptide of the TAP1(RING4) sequence²⁰. This reagent precipitated the TAP1 protein and coprecipitated the associated TAP2 product²⁰. Preliminary cell staining experiments were done to compare the mutant lymphoblastoid cell line LCL721.174 (termed .174), that has deleted both TAP genes^{9,17}, with the wild-type control LCL721 (termed 721). The unpurified serum gave high background staining on .174 cells (data not shown; ref 20), but after affinity purification on the TAP1 peptide used for immunization, it stained 721 cells specifically (Fig. 1a). The improved specificity of staining after affinity purification was accompanied by the

TABLE 1 Numbers of anti-TAP1 immunogold particles on cryosections of LCL721 and .174 cells

	LCL721 cells	.174 cells	Specificity factor
Rough endoplasmic reticulum	671	96	7.0
Golgi-complex	59	22	2.7
Irrelevant*	696	845	0.8
Total gold	1,426	963	

Cells were randomly selected in the electron microscope viewer and were photographed at a magnification of $\times 3000$ until a section surface of $1.815 \mu\text{m}^2$ had been included for each cell type. In these cell profiles gold particles were counted and assigned to the structures listed. A gold particle was considered to label a membrane when lying within 20 nm of it. Totals of 19 LCL 721 cells and 24 mutant .174 cells were quantified. Of each cell type $\sim 180 \mu\text{m}^2$ endoplasmic reticulum and $40 \mu\text{m}^2$ Golgi complex were evaluated. The specificity factor of TAP1 labelling is the ratio of LCL721 over .174 labelling for a particular compartment.

* 'Irrelevant', for this study included nuclei, mitochondria, cytosol and unidentifiable structures.

loss of two prominent proteins of $M_r \sim 33$ – 34 K in immunoprecipitates from both .174 and 721 cells (Fig. 1b), which may have arisen through cross-reactions with the carrier (KLH) used for immunization. Additional proteins of ~ 28 K and ~ 52 K were precipitated from the wild-type 721 cells but not from .174, and were retained after affinity purification of the AK1.7 serum. We are investigating the nature of these bands. They may have co-precipitated with the TAP1 protein, because western blot analysis with the affinity purified AK1.7 antiserum detected a single ~ 71 K band in total cell lysates from 721 cells, which was absent in similar extracts from .174 (Fig. 1c).

The distribution of staining revealed by light microscopy on acetone-fixed cytospin preparations was predominantly perinuclear, consistent with location in the endoplasmic reticulum (Fig. 1a). To determine the precise location of the TAP1 protein we performed high resolution immuno-electron microscopy on ultrathin cryosections of 721 and .174 cells. Immunogold particles were counted on sections of the two cell lines and were found specifically on the membranes of the endoplasmic reticulum and Golgi complex (Figs 2, 3a, b and Table 1). Because the antibody was raised to a peptide from the C terminus of the TAP1 ATP-binding domain and labelling occurred predominantly on the cytosolic side of the membrane (Fig. 2) we suggest that the TAP1 protein is oriented in the endoplasmic reticulum and Golgi membranes with its ATP-binding domain in the cytosol.

The distribution of TAP1 in the Golgi complex was defined by comparison to class I MHC molecules (which can be detected throughout the stack)²³, and p53 (which is localized to the *cis*-Golgi)^{24,25}. Although MHC class I molecules were found throughout the stacks of Golgi cisternae, TAP1 was confined to only one side of the stacks (Fig. 3a). Comparison with p53 staining showed that TAP1 was localized to the *cis*-Golgi (Fig. 3b). We estimate that the density of TAP staining of the *cis*-Golgi is close to that of the endoplasmic reticulum, because the *cis*-Golgi cisternae comprises only $\sim 1/4$ to $\sim 1/3$ of the total Golgi membranes. The specificity factor for *cis*-Golgi, as opposed to total Golgi (in Table 1), would therefore be ~ 9 . Expression of the TAP complex in the *cis*-Golgi as well as the endoplasmic reticulum may provide a hint as to the sites of peptide loading. In fact, it has been suggested that class I molecules cycle between the endoplasmic reticulum and *cis*-Golgi⁴¹.

The TAP proteins are the only members of the ABC superfamily of transporters that localize to the endoplasmic reticulum and *cis*-Golgi. The mechanism of localization of the TAP1 protein is not apparent, because it does not contain any previously characterized endoplasmic reticulum retention signals^{26,27}. To see if retention of TAP1 required its association with TAP2, we stained mutant T2 cells that expressed the TAP1

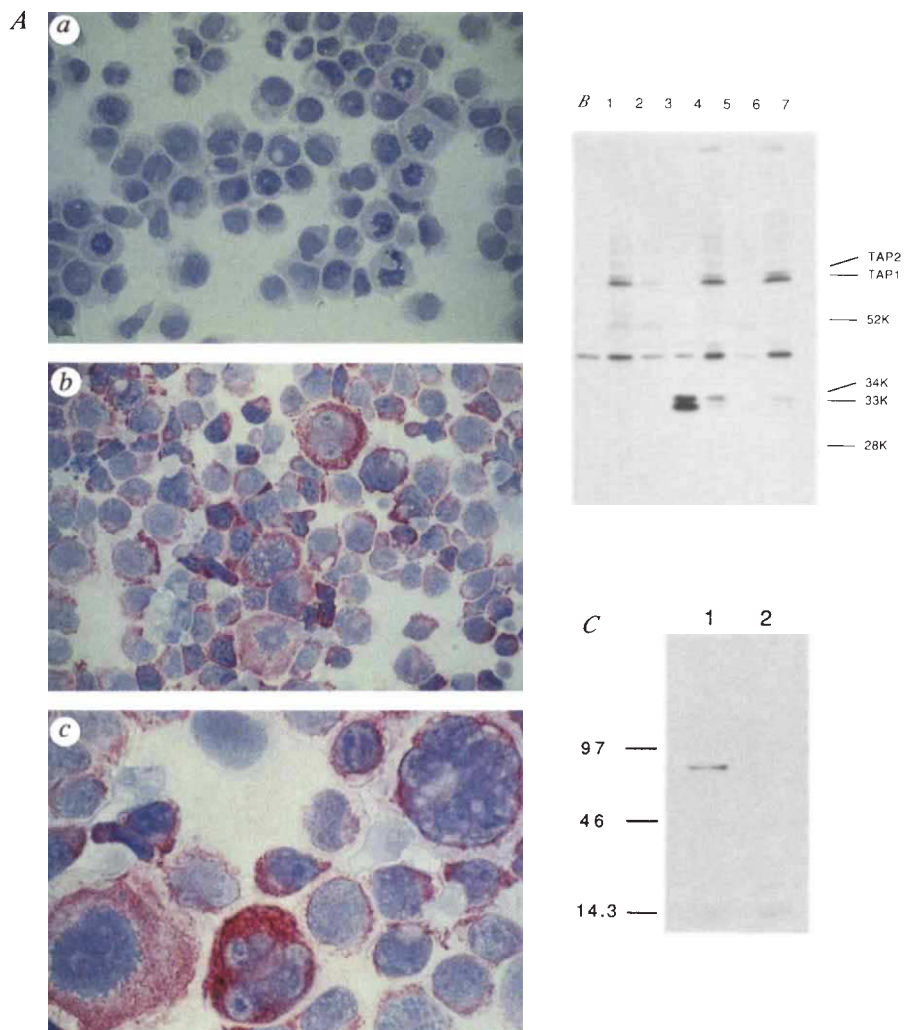
protein in isolation²⁰. The T2 cell line was derived from .174 and carries the same MHC deletion¹⁰. The distribution of TAP1 in the transfected T2 cells was identical to that of the non mutant 721 (Fig. 3c). This result demonstrated that localization of the TAP1 protein required neither the formation of the TAP1/TAP2

complex nor the expression of other genes in the .174 deletion, including the LMP2 and LMP7 proteasome components^{28,29}.

The loss of peptide antigen presentation and class I molecule stability in the mutant cell lines BM36.1, RMA-S and .134 results from defects in the TAP1/TAP2 complex^{20,21,30}. The instability

FIG. 1 A, Staining of .174 (a) and 721 cells (b, c) with affinity-purified AK1.7 antibody. (c, High magnification) B, Immunoprecipitation. The lanes represent extracts from cell lines precipitated with the following antibody preparations: lane 1, .174 mutant, 15 μ g affinity purified AK1.7; 2, 721 wild-type cells, 15 μ g affinity-purified AK1.7; 3, 721 cells, 5 μ g affinity purified AK1.7; 4, .174 cells, 1/40 AK1.7 serum; 5, 721 cells, 1/40 AK1.7 serum; 6, 721 cells, 15 μ g IgG fraction of AK1.7 serum; 7, 721 cells, 1/40 AK1.6 serum. The ~71K TAP1 and ~77K TAP2 proteins are labelled. Also indicated are the 33–34K doublet that is removed by affinity purification and the ~52 and ~28K proteins that are not (see text). C, Western blotting of Nonidet-P40 lysates of 721 (lane 1) and .174 (lane 2) cells probed with the affinity purified AK1.7 anti-TAP1 antiserum.

METHODS. Affinity purification of AK1.7 serum. The serum was obtained as described²⁰ after the sixth injection of antigen into a Sandy half-lop rabbit. The TAP1 peptide GCYWAMVQAPADAPE (single-letter amino-acid code) was linked directly to CNBr-activated Sepharose 4B (Pharmacia). Dry gel (0.149 g) was activated in 50 ml 1 mM HCl for 5 min, washed once in 1 mM HCl and resuspended in 20 ml 0.1 M NaHCO₃ containing 5.4 mg dissolved peptide. After 2 h incubation at room temperature ~40% of peptide had coupled (5.4 mg peptide per ml swollen gel). The gel was washed in 0.1 M Tris-HCl pH8, packed into a column (0.4 ml) and washed with 4 M MgCl₂ followed by 0.1 M TrisHCl pH8. Filtered serum (8 ml) was passed continuously over the column for 45 min. The column was then washed with 0.1 M Tris-HCl pH8 until the optical density at 280 nm (OD280) of the fall through was 0.014. Bound antibody was eluted with 4 M MgCl₂. Positive fractions were pooled and dialysed against Tris buffered saline pH8. This material was further affinity purified by elution at pH3 from protein A Sepharose³⁸ and 0.22 μ m filtered. The final yield was 119 μ g affinity-purified antibody per ml serum, which represented ~2.8% of the available IgG. Indirect immuno-alkaline phosphatase staining (APAP) of acetone-fixed cytosin preparations of 721 and .174 and immunoprecipitation of the TAP1/TAP2 complex were done as described^{39,20}. For immunoprecipitation, sera AK1.6 (taken after the fifth immunization) and AK1.7 (after the sixth) were added to 1/40 dilution, purified antibody to 15 μ g ml⁻¹. For western blots, total cell lysates were prepared from 5 $\times$ 10⁴ cells and separated by SDS-PAGE



in reducing sample buffer before transfer by electroblotting to Hybond C Extra nitrocellulose membranes. Blots were incubated with 1/1,000 (approx. 133 ng/ml) dilutions of the AK1.7 antiserum as the primary antibody and 1/1,000 dilution of horseradish peroxidase-conjugated swine anti-rabbit IgG as the second antibody. Detection of bands was by enhanced chemiluminescence (Amersham).

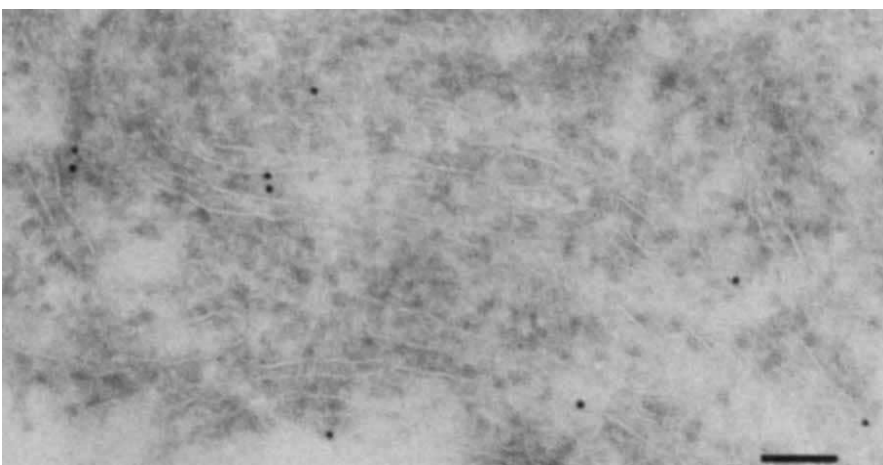


FIG. 2 Electron micrograph of an ultrathin cryosection of a 721 cell, immuno-labelled with affinity purified AK1.7. Note the labelling at the cytosolic side of rough endoplasmic reticulum. Scale bar, 125 nm.

METHODS. Cells were fixed with 2% glutaraldehyde in 0.1 M phosphate buffer at pH 7.4 for 2 h at room temperature. Ultrathin cryosectioning and immunogold labelling were done as described⁴⁰. Briefly, after fixation cells were washed twice with PBS buffer and twice with PBS, 0.15 M glycine. After embedding in 10% gelatin the slab was kept at 4 °C. Small pieces of gelatin were infiltrated with 2.3 M sucrose, mounted and frozen in liquid nitrogen. Ultrathin cryosections were immuno-labelled with affinity-purified rabbit anti-TAP1 IgG, swine anti-rabbit IgG and finally labelled with 10 nm protein A gold particles.

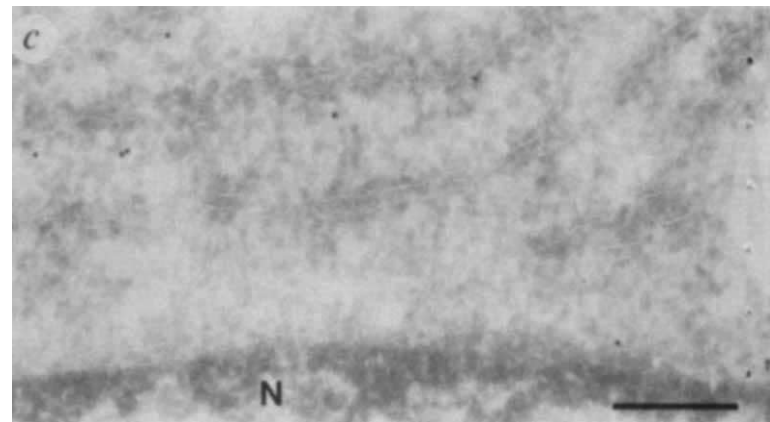
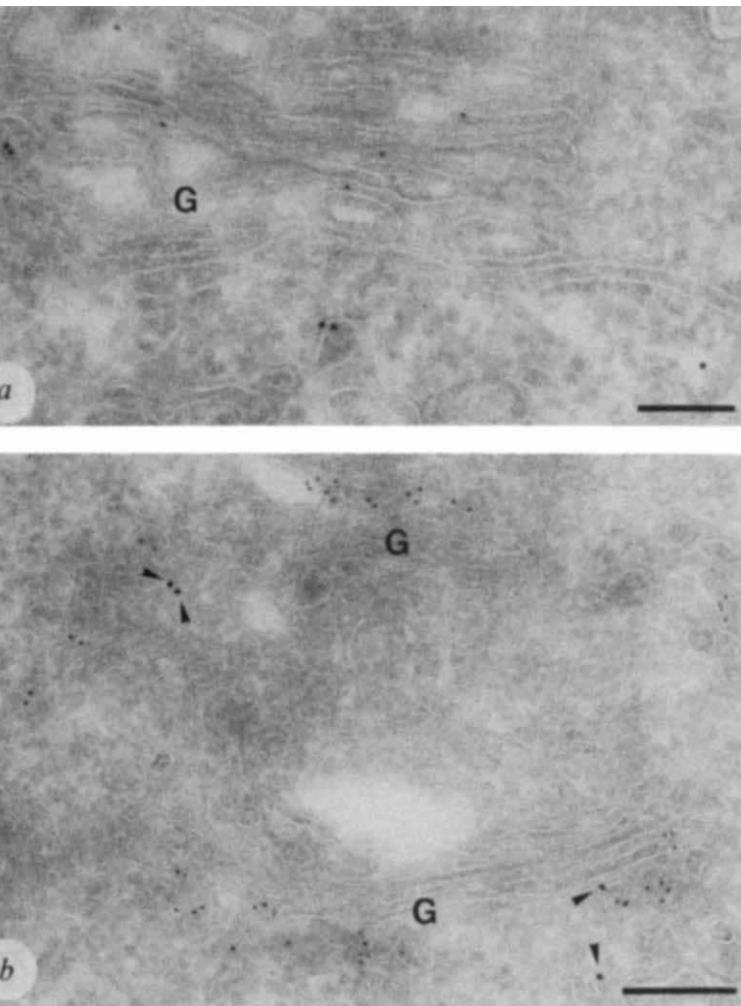


FIG. 3 *a*, Section of 721 cells double-labelled with small gold for heavy chain of MHC class I molecules and with large gold particles for TAP1. Although class I heavy chain occurs throughout the Golgi stack of cisternae (G), TAP1 is confined to one side of the stack. Scale bar, 125 nm. *b*, The TAP1 (large gold, arrowheads) co-localizes with the *cis*-Golgi marker p53 (small gold). G, Golgi stacks. Scale bar, 200 nm. *c*, TAP1 labelling at the rough endoplasmic reticulum and perinuclear cisternae in a T2 cell transfected with TAP1 complementary DNA. N, nucleus. Scale bar, 250 nm.

METHODS. In the double-labelling procedure the sections were first labelled with the TAP1 antibody (as for Fig. 2) followed by either rabbit anti-MHC class I heavy chain serum or the monoclonal antibody mouse-anti p53. In the case of p53, rabbit anti-mouse IgG1 was used as a bridging step and both second antibodies were labelled with 5 nm protein A-gold particles.

of class I molecules from the mutants is reversed by binding peptides *in vitro*^{11–13,31}. In normal cells, class I heavy chains usually form a stable complex with β_2 microglobulin within minutes of synthesis³². At this time they retain sensitivity to endoglycosidase H, implying that the stable, peptide containing complex forms in the endoplasmic reticulum. The inhibitory effects of brefeldin A on presentation of intracellular antigens also indicated that peptides associate with class I molecules in the endoplasmic reticulum or *cis*-medial Golgi^{33,34}. The location of the TAP complex in the endoplasmic reticulum and *cis*-Golgi is therefore consistent with its proposed role in transporting

peptides from the cytosol to the endoplasmic reticulum in living cells^{15–18}. The mutation in the TAP2 ATP-binding domain of mutant BM36.1 indicates that this function may be ATP dependent²⁰.

Recent experiments show that *in vitro*, microsomes prepared from T2 cells (that lack TAP proteins) are freely permeable to peptides in the absence of ATP, but are defective in class I molecule assembly³⁵. This has led to the suggestion that the primary defect in T2 is not in peptide transport into the endoplasmic reticulum, but in the assembly of class I heavy chains with peptide and β_2 -microglobulin^{35,36}. But these data are not consistent with experiments on intact cells from a variety of mutants which are more easily explained by a peptide transport defect^{11–13,20,22,31,37}. These issues will not be fully resolved until the function of the TAP complex can be studied in isolation. Antibodies with the properties we have described may be helpful in the purification and analysis of the TAP complex. □

Received 13 February; accepted 22 April 1992.

1. Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27**, 51–177 (1979).
2. Townsend, A. R. M., Gotch, F. M. & Davey, J. *Cell* **42**, 457–467 (1985).
3. Townsend, A. R. M. *et al. Cell* **44**, 959–968 (1986).
4. Bjorkman, P. J. *et al. Nature* **329**, 512–518 (1987).
5. Townsend, A. & Bodmer, H. A. *Rev. Immun.* **7**, 601–24 (1989).
6. Yewdell, J. W. & Bennink, J. R. *Cell* **62**, 203–206 (1990).
7. Braciale, T. J. & Braciale, V. L. *Immun. Today* **12**, 124–129 (1991).
8. Ljunggren, H. G. *et al. J. Immun.* **142**, 2911–2917 (1989).
9. DeMars, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 8183–8187 (1985).
10. Salter, R. D. & Cresswell, P. *EMBO J.* **5**, 943–949 (1986).
11. Townsend, A. *et al. Nature* **340**, 443–448 (1989).
12. Townsend, A. *et al. Cell* **62**, 285–295 (1990).
13. Cerundolo, V. *et al. Nature* **345**, 449–452 (1990).
14. Hosken, N. A. & Bevan, M. J. *Science* **248**, 367–370 (1990).
15. Deverson, E. V., *et al. Nature* **348**, 738–741 (1990).
16. Trowsdale, J. *et al. Nature* **348**, 741–744 (1990).
17. Spies, T. *et al. Nature* **348**, 744–747 (1990).
18. Monaco, J. J., Cho, S. & Attaya, M. *Science* **250**, 1723–1726 (1990).
19. Powis, S. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 1463–1467 (1992).
20. Kelly, A. *et al. Nature* **355**, 641–644 (1992).
21. Powis, S. J. *et al. Nature* **354**, 528–531 (1991).
22. Spies, T. *et al. Nature* **355**, 644–646 (1992).

23. Neefjes, J. J., Stollorz, V., Peters, P. J., Geuze, H. J. & Pleogh, H. L. *Cell* **61**, 171–183 (1990).
24. Schweizer, A., Franssen, J. A. M., Baechli, T., Ginsel, L. & Hauri, H. P. *J. Cell Biol.* **107**, 1643–1653 (1988).
25. Schweizer, A., Matter, K., Ketcham, C. M. & Hauri, H. P. *J. Cell Biology* **113**, 45–54 (1991).
26. Nilsson, T., Jackson, M. & Peterson, P. A. *Cell* **58**, 707–718 (1989).
27. Pelham, H. R. B. *Trends biochem. Sci.* **15**, 483–486 (1990).
28. Glynn, R. *et al. Nature* **353**, 357–360 (1991).
29. Kelly, A. *et al. Nature* **353**, 667–668 (1991).
30. Spies, T. & DeMars, R. *Nature* **351**, 323–324 (1991).
31. Schumacher, T. N. M. *et al. Cell* **62**, 563–567 (1990).
32. Owen, M. J., Kissner, A. M. & Lodish, H. F. *J. Biol. Chem.* **255**, 9678–9684 (1980).
33. Nuchtern, J. G., Bonifacio, J. S., Biddison, W. E. & Klausner, R. D. *Nature* **339**, 223–226 (1989).
34. Yewdell, J. W. & Bennink, J. R. *Science* **244**, 1072–1075 (1989).
35. Levy, F., Gabathuler, R., Larsson, R. & Kvist, S. *Cell* **67**, 265–274 (1991).
36. Dobberstein, B. *Nature* **355**, 109–110 (1992).
37. Anderson, K. *et al. J. exp. Med.* **174**, 489–492 (1991).
38. Ey, P. L., Prowse, S. J. & Jenkin, C. R. *Immunochimistry* **15**, 429–436 (1978).
39. Cordell, J. L. *et al. J. Histochem. Cytochem.* **32**, 219–229 (1984).
40. Slot, J. W., Geuze, H. J. & Weerkamp, A. J. *Meth. Microbiol.* **20**, 211 (1988).
41. Hsu, V. W. *et al. Nature* **352**, 441–444 (1991).

ACKNOWLEDGEMENTS We thank M. Jones and D. Mason for APAP staining cells as well as E. Souter and G. Warren for advice. We also thank K. Micklem for advice on affinity purification, B. DeMars for providing LCL721 and 174 cell lines and the MRC and Cancer Research Institute (NY) for support.

Spectroscopic methods

New products in the spectroscopic arena include an advanced spectral library system for infrared spectroscopy and a fast-scanning UV/visible spectrophotometer that features reverse optics.

3M has introduced a new **disposable sample card** that is designed to replace traditional salt plates for many infrared spectroscopy applications (*Reader Service No. 101*). Measuring 5×10 cm, the 3M brand disposable infrared card conforms to most FTIR spectrometers. A 2-cm aperture in the centre of the card exposes a thin, microporous substrate. Samples are applied directly to the substrate. 3M says its porosity speeds evaporation and drying. And, unlike traditional salt plates, which must be



Disposable IR cards for FTIR specs.

carefully cleaned and polished and which are prone to break, the 3M infrared card requires very little preparation time. There is no need for cleaning solvents or tissues, reducing the time analysts are exposed to hazardous samples and solvents, 3M says. It is also intended to reduce the chance of sample cross-contamination. The card can be used for analysing organic liquids, substances that are soluble in organic solvents, and semi-solids.

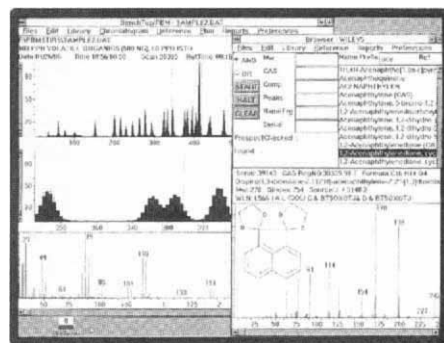
The Advanced Spectral Library System (ASLS) for infrared spectroscopy from Lab Control represents a new **laboratory integration and management system** designed for modern laboratories (*Reader Service No. 102*). The system takes full advantage of the recent developments in infrared and Raman spectroscopy, as well as in the field of computer hardware and software. The ASLS integrates all aspects of infrared and Raman spectroscopy, beginning with data acquisition on the individual spectrometer, subsequent standard and advanced qualitative and quantitative evaluation of the acquired data, including final reporting, documentation and archiving of the evaluated spectral data. The system facilitates the generation and administration of centralized data collections and enables an easy data exchange between different computer and spectrometer systems. The ASLS is designed to work in a local and wide area

network environment built up from DEC VAX computers and PCs. Lab Control also offers version 2.0 of Spectrafile-IR *plus*, a program for online data acquisition, mathematical manipulation, quantitative evaluation, library searching, documentation and data conversion for infrared spectroscopy. The program offers mathematical functions, including Kramers-Kronig-Kubelka-Munk transformation and Fourier self deconvolution. It supports all functions of FTIR and dispersive instruments, modern and old spectrometers as well, providing all the features that are needed to set up and control infrared spectrometers. Data collection from spectrometers without a computer interface can be controlled with A/D converter boards. The software is available in English, German and French; Spanish and Italian versions will be available in autumn.

Interpretation of data

SpecInfo, originally developed by BASF to speed up the interpretation of NMR, IR and MS spectra, is now marketed independently by Chemical Concepts. A new version, based on a relational database system, has just been released (*Reader Service No. 103*). The use of a relational database offers several advantages in data handling. For example, security and data integrity are easily managed with two levels of access available. The company says that entry of structural and spectral data is now much easier and the program is able to import a wide range of file formats, including JCAMP-DX and NMR peaklists for spectra and Molecular Design Limited's 'Molfile' format for structures. Besides normal library functions such as searching for similar or identical structures and spectra, SpecInfo offers a range of tools to help the researcher in the elucidation of unknown structures. The system can search for individual spectral features, then analyse the results and suggest substructures that correspond to these features. Thus, the researcher builds up a list of likely substructures, like jigsaw pieces, which can then be assembled into a plausible molecular structure. However, Chemical Concepts says that SpecInfo goes a step further. Once the jigsaw of substructures has been assembled into a molecule, SpecInfo is able to calculate its spectrum. By comparing this with the measured spectrum, it should be easy to see if the proposed structure is correct.

Palisade announces the release of BenchTop/PBM 2.2. This new version is a

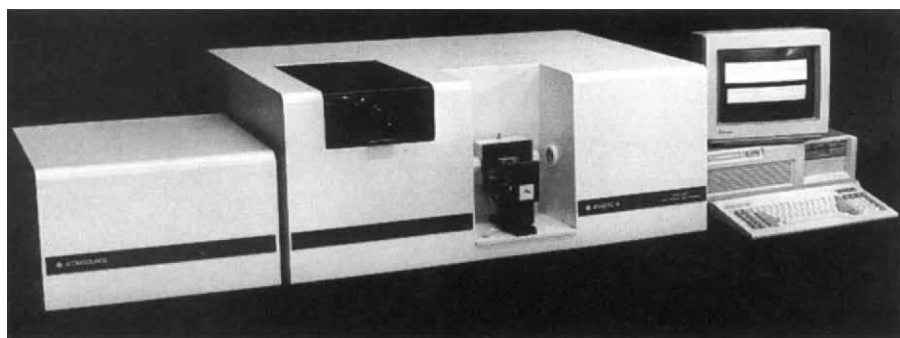


Mass spectral library search system.

major update to the **PC-based mass spectral library search system**, first released in 1989 (*Reader Service No. 104*). BenchTop/PBM uses the Windows environment to combine PBM library search algorithms with the Wiley Registry of Mass Spectral Data (5th edition). Version 2.2 includes extensive user library capabilities, Windows 3.0 support, expanded background subtraction features, a new mass chromatogram display, new features for editing and averaging scans, additional file format support, and a variety of other new user interface features. Multiple proprietary libraries can be added to the BenchTop/PBM to supplement or replace the Wiley Registry. Any number of new libraries can be created of any desired size. Additional spectra can be added to an existing user library at any time. Copies of BenchTop/PBM 2.2 cost \$3,995 (US). This price includes the Wiley Registry of Mass Spectral Data.

High-tech hardware

Analyte has introduced a **furnace atomizer** for its Model 16 atomic absorption spectrophotometer, which allows the system to determine up to 24 elements in rapid sequence in an individual sample, with each element being measured under optimum conditions (*Reader Service No. 105*). With this recent addition, the instrument can be fitted with three different atomizers: a flame atomizer for liquids, an atomsource sputtered atomization device for the direct analysis of solids, and the new furnace for measuring very low concentrations. Most multi-element atomic absorption spectrophotometers run all the samples for one element, then switch to another. Consequently, Analyte says, eight elements are not completed in the first sample until, say, 30 samples have been analysed for seven. And, whereas some furnace atomic absorption spectrophotometers are designed to

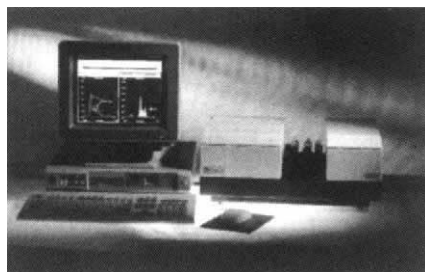


Analyte 16 — can determine the concentration of up to 24 elements in rapid sequence.

measure up to four elements simultaneously, Analyte says that often a compromise must be made between the pyrolysis and atomization times and temperatures of the four elements. The Analyte 16 avoids these problems by analysing an individual sample for all its required elements in sequence, with each element run at its optimum times and temperatures. In other words, the first sample is completed before the second sample is started.

Reverse optics — where light is monochromated post sample — are a feature of the new Biochrom 4060 fast-scanning UV/visible spectrophotometer from Pharmacia Biotech (*Reader Service No. 106*). This advanced optical configuration is designed to provide maximum detector stability and high resolution. The open sample compartment allows for sample handling during kinetic runs. Samples can be added and

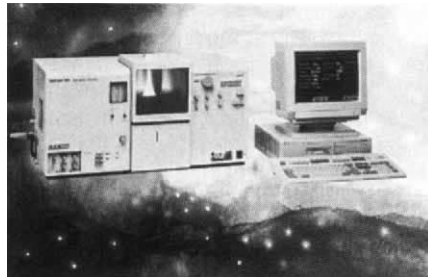
mixed immediately before or during data acquisition. A range of sampling accessories is also available to enhance further the flexibility of the system. Biochrom 4060's Microsoft Windows-based software features real-time visual display during data acquisition, three-dimensional overlay, post-run mathematical manipulation and mouse-driven parameter selection. Stated applications for the 4060 include: wavelength scanning at $2,400 \text{ nm min}^{-1}$, multiple-wavelength analysis including DNA quan-



Pharmacia's fast-scanning UV/visible spec features reverse optics.

tification and purity determination, one or two point reaction kinetics with various Michaelis-Menten plots, time drive for online analysis, quantification and fraction analysis.

Unicam has added the Solaar 919 model to the line up of **atomic absorption spectrometers** in its Solaar series (*Reader Service No. 107*). The instrument features high energy optics and an advanced universal spray chamber designed to provide good flame sensitivity. Unicam says that trace analysis performance is assured by the addition of the new GF90 furnace and autoprobe system. Dynamic optical temperature control is intended to provide reproducible



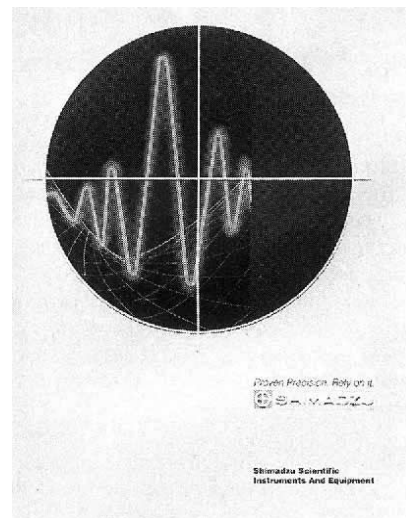
Solaar 919 — designed as a cost-effective atomic absorption solution.

heating rates of over $2,000^\circ\text{C}$. Furnace autoprobe ensures the most effective control of vapour phase interferences. The modular design of the Solaar 919, should facilitate the addition of a quad lamp turret, advanced gas control and automatic background correction. The optional data station provides methods and results storage and uses an intuitive user interface with pop-up menus and graphics.

Catalogues

Over 300 different types of cell, most of which are available in a choice of optical glass, UV, quartz, infrared quartz or ES quartz, are featured in the 1992 **spectrophotometer cell catalogue** from NSG Precision Cells (*Reader Service No. 108*). The catalogue includes a selection of standard cells, anaerobic cells, micro cells, water-jacketed cells and dye laser cells. NSG also features accessories such as cell washers and mixers.

A range of instruments for chromatography, spectroscopy, environmental and physical testing is featured in Shimadzu's 32-page **catalogue of analytical instrumentation** (*Reader Service No. 109*). In addition to the LC-10A Series HPLC



Shimadzu's Instrumentation catalogue.

system, the catalogue features gas chromatography, PC-based chromatography data systems, integrators, UV/visible, fluorescence, atomic absorption and infrared spectrophotometers, scanning densitometers, thermal analysers, total organic analysers, particle size analysers and arc-spark emission spectrometers. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.

ADVERTISEMENTS

Highest Quality Oligonucleotides

Synthesised, purified and shipped in 3 days

£2.75 per base

(All inclusive)

0.2 μM scale (approx. 500 μg)

- Antisense (completely or partially sulphurised)
- Extensive range of labelling and modifications
- Full design and analysis facilities
- Gene cassettes, bulk order and contract discounts

Molecular Medicine Unit

King's College School of Medicine & Dentistry

Tel. 071 - 326 3126 FAX 071 - 733 3877

Reader Service No.56

Synthetic Oligonucleotides

Rapid Quality Worldwide Service

£2.30 per base, no setup!

(0.2 μmole - limited offer)

option of H.P.L.C. purification - £30

biotinylation, fluorescent labelling

phosphothioates, many other modifications

OSWEL DNA SERVICE

Department of Chemistry, University of Edinburgh
Kings Buildings, West Mains Road,
EDINBURGH EH9 3JJ

Tel: 031 650 4792 Fax: 031 662 4054

Reader Service No.9